

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/31, C07K 14/22, 16/12, A61K 39/095 // (C12N 15/31, C12R 1:36)</b>		A2	(11) International Publication Number: <b>WO 99/24578</b>
			(43) International Publication Date: 20 May 1999 (20.05.99)
(21) International Application Number: PCT/IB98/01665		(74) Agent: HALLYBONE, Huw, George; Carpmals & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).	
(22) International Filing Date: 9 October 1998 (09.10.98)		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BI, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(30) Priority Data: 9723516.2 6 November 1997 (06.11.97) GB 9724190.5 14 November 1997 (14.11.97) GB 9724386.9 18 November 1997 (18.11.97) GB 9725158.1 27 November 1997 (27.11.97) GB 9726147.3 10 December 1997 (10.12.97) GB 9800759.4 14 January 1998 (14.01.98) GB 9819016.8 1 September 1998 (01.09.98) GB		Published Without international search report and to be republished upon receipt of that report.	
(71) Applicant (for all designated States except US): CHIRON S.P.A. [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT).			
(72) Inventors; and (75) Inventors/Applicants (for US only): MASIGNANI, Vega [IT/IT]; Via Pantaneto, 105, I-53100 Siena (IT). RAP-PUOLI, Rino [IT/IT]; Via delle Rocche, 1, Vagliagli, I-53019 Castelnuovo Berardenga (IT). PIZZA, Mariagrazia [IT/IT]; Strada di Montalbuccio, 160, I-53100 Siena (IT). SCARLATO, Vincenzo [IT/IT]; Via Firenze, 3/37, I-53134 Colle Val d'Elsa (IT). GRANDI, Guido [IT/IT]; 9° Strada, 4, I-20090 Segrate (IT).			
(54) Title: NEISSERIAL ANTIGENS			
(57) Abstract <p>The invention provides proteins from <i>Neisseria meningitidis</i> (strains A and B) and from <i>Neisseria gonorrhoeae</i> including amino acid sequences, the corresponding nucleotide sequences, expression data, and serological data. The proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		



## NEISSERIAL ANTIGENS

This invention relates to antigens from *Neisseria* bacteria.

### BACKGROUND ART

*Neisseria meningitidis* and *Neisseria gonorrhoeae* are non-motile, gram negative diplococci that are pathogenic in humans. *N.meningitidis* colonises the pharynx and causes meningitis (and, occasionally, septicaemia in the absence of meningitis); *N.gonorrhoeae* colonises the genital tract and causes gonorrhea. Although colonising different areas of the body and causing completely different diseases, the two pathogens are closely related, although one feature that clearly differentiates meningococcus from gonococcus is the presence of a polysaccharide capsule that is present in all pathogenic meningococci.

*N.gonorrhoeae* caused approximately 800,000 cases per year during the period 1983-1990 in the United States alone (chapter by Meitzner & Cohen, "Vaccines Against Gonococcal Infection", In: *New Generation Vaccines*, 2nd edition, ed. Levine, Woodrow, Kaper, & Cobon, Marcel Dekker, New York, 1997, pp.817-842). The disease causes significant morbidity but limited mortality. Vaccination against *N.gonorrhoeae* would be highly desirable, but repeated attempts have failed. The main candidate antigens for this vaccine are surface-exposed proteins such as pili, porins, opacity-associated proteins (Opas) and other surface-exposed proteins such as the Lip, Laz, IgA1 protease and transferrin-binding proteins. The lipooligosaccharide (LOS) has also been suggested as vaccine (Meitzner & Cohen, *supra*).

*N.meningitidis* causes both endemic and epidemic disease. In the United States the attack rate is 0.6-1 per 100,000 persons per year, and it can be much greater during outbreaks (see Lieberman *et al.* (1996) Safety and Immunogenicity of a Serogroups A/C *Neisseria meningitidis* Oligosaccharide-Protein Conjugate Vaccine in Young Children. *JAMA* 275(19):1499-1503; Schuchat *et al* (1997) Bacterial Meningitis in the United States in 1995. *N Engl J Med* 337(14):970-976). In developing countries, endemic disease rates are much higher and during epidemics incidence rates can reach 500 cases per 100,000 persons per year. Mortality is extremely high, at 10-20% in the United States, and much higher in developing countries. Following the introduction of the conjugate vaccine against *Haemophilus influenzae*, *N. meningitidis* is the major cause of bacterial meningitis at all ages in the United States (Schuchat *et al* (1997) *supra*).

Based on the organism's capsular polysaccharide, 12 serogroups of *N.meningitidis* have been identified. Group A is the pathogen most often implicated in epidemic disease in sub-Saharan Africa. Serogroups B and C are responsible for the vast majority of cases in the United States and in most developed countries. Serogroups W135 and Y are responsible for the rest of the cases in the United States and developed countries. The meningococcal vaccine currently in use is a tetravalent polysaccharide vaccine composed of serogroups A, C, Y and W135. Although efficacious in adolescents and adults, it induces a poor immune response and short duration of protection, and cannot be used in infants [eg. Morbidity and Mortality weekly report, Vol.46, No. RR-5 (1997)]. This is because polysaccharides are T-cell independent antigens that induce a weak immune response that cannot be boosted by repeated immunization. Following the success of the vaccination against *H.influenzae*, conjugate vaccines against serogroups A and C have been developed and are at the final stage of clinical testing (Zollinger WD "New and Improved Vaccines Against Meningococcal Disease" in: *New Generation Vaccines*, supra, pp. 469-488; Lieberman *et al* (1996) supra; Costantino *et al* (1992) Development and phase I clinical testing of a conjugate vaccine against meningococcus A and C. *Vaccine* 10:691-698).

Meningococcus B remains a problem, however. This serotype currently is responsible for approximately 50% of total meningitis in the United States, Europe, and South America. The polysaccharide approach cannot be used because the menB capsular polysaccharide is a polymer of  $\alpha(2-8)$ -linked *N*-acetyl neuraminic acid that is also present in mammalian tissue. This results in tolerance to the antigen; indeed, if an immune response were elicited, it would be anti-self, and therefore undesirable. In order to avoid induction of autoimmunity and to induce a protective immune response, the capsular polysaccharide has, for instance, been chemically modified substituting the *N*-acetyl groups with *N*-propionyl groups, leaving the specific antigenicity unaltered (Romero & Outschoorn (1994) Current status of Meningococcal group B vaccine candidates: capsular or non-capsular? *Clin Microbiol Rev* 7(4):559-575).

Alternative approaches to menB vaccines have used complex mixtures of outer membrane proteins (OMPs), containing either the OMPs alone, or OMPs enriched in porins, or deleted of the class 4 OMPs that are believed to induce antibodies that block bactericidal activity. This approach produces vaccines that are not well characterized. They are able to protect against the homologous strain, but are not effective at large where there are many antigenic variants of the outer membrane proteins. To overcome the antigenic variability, multivalent vaccines containing up to nine different

porins have been constructed (eg. Poolman JT (1992) Development of a meningococcal vaccine. *Infect. Agents Dis.* 4:13-28). Additional proteins to be used in outer membrane vaccines have been the opa and opc proteins, but none of these approaches have been able to overcome the antigenic variability (eg. Ala' Aldeen & Borriello (1996) The meningococcal transferrin-binding proteins 1 and 2 are both surface exposed and generate bactericidal antibodies capable of killing homologous and heterologous strains. *Vaccine* 14(1):49-53).

A certain amount of sequence data is available for meningococcal and gonococcal genes and proteins (eg. EP-A-0467714, WO96/29412), but this is by no means complete. The provision of further sequences could provide an opportunity to identify secreted or surface-exposed proteins that are presumed targets for the immune system and which are not antigenically variable. For instance, some of the identified proteins could be components of efficacious vaccines against meningococcus B, some could be components of vaccines against all meningococcal serotypes, and others could be components of vaccines against all pathogenic *Neisseriae*.

## THE INVENTION

The invention provides proteins comprising the Neisserial amino acid sequences disclosed in the examples. These sequences relate to *N.meningitidis* or *N.gonorrhoeae*.

It also provides proteins comprising sequences homologous (ie. having sequence identity) to the Neisserial amino acid sequences disclosed in the examples. Depending on the particular sequence, the degree of identity is preferably greater than 50% (eg. 65%, 80%, 90%, or more). These homologous proteins include mutants and allelic variants of the sequences disclosed in the examples. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between the proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

The invention further provides proteins comprising fragments of the Neisserial amino acid sequences disclosed in the examples. The fragments should comprise at least *n* consecutive amino acids from the sequences and, depending on the particular sequence, *n* is 7 or more (eg. 8, 10, 12, 14, 16, 18, 20 or more). Preferably the fragments comprise an epitope from the sequence.

The proteins of the invention can, of course, be prepared by various means (*eg.* recombinant expression, purification from cell culture, chemical synthesis *etc.*) and in various forms (*eg.* native, fusions *etc.*). They are preferably prepared in substantially pure or isolated form (*ie.* substantially free from other Neisserial or host cell proteins)

- 5 According to a further aspect, the invention provides antibodies which bind to these proteins. These may be polyclonal or monoclonal and may be produced by any suitable means.

According to a further aspect, the invention provides nucleic acid comprising the Neisserial nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences homologous (*ie.* having sequence identity) to the Neisserial nucleotide  
10 sequences disclosed in the examples.

Furthermore, the invention provides nucleic acid which can hybridise to the Neisserial nucleic acid disclosed in the examples, preferably under "high stringency" conditions (*eg.* 65°C in a 0.1xSSC, 0.5% SDS solution).

Nucleic acid comprising fragments of these sequences are also provided. These should comprise  
15 at least  $n$  consecutive nucleotides from the Neisserial sequences and, depending on the particular sequence,  $n$  is 10 or more (*eg.* 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

It should also be appreciated that the invention provides nucleic acid comprising sequences  
20 complementary to those described above (*eg.* for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (*eg.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*eg.* single stranded, double stranded, vectors, probes *etc.*).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as  
25 those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (eg. expression vectors) and host cells transformed with such vectors.

According to a further aspect, the invention provides compositions comprising protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as vaccines,  
5 for instance, or as diagnostic reagents, or as immunogenic compositions.

The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (eg. as vaccines) or as diagnostic reagents. It also provides the use of nucleic acid, protein, or antibody according to the invention in the manufacture of: (i) a medicament for treating or preventing infection due to Neisserial bacteria; (ii) a diagnostic reagent for detecting the  
10 presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria; and/or (iii) a reagent which can raise antibodies against Neisserial bacteria. Said Neisserial bacteria may be any species or strain (such as *N.gonorrhoeae*, or any strain of *N.meningitidis*, such as strain A, strain B or strain C).

The invention also provides a method of treating a patient, comprising administering to the patient  
15 a therapeutically effective amount of nucleic acid, protein, and/or antibody according to the invention.

According to further aspects, the invention provides various processes.

A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

20 A process for producing protein or nucleic acid of the invention is provided, wherein the the protein or nucleic acid is synthesised in part or in whole using chemical means.

A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting said duplexes.

25 A process for detecting proteins of the invention is provided, comprising the steps of: (a) contacting an antibody according to the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

A summary of standard techniques and procedures which may be employed in order to perform the invention (eg. to utilise the disclosed sequences for vaccination or diagnostic purposes) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

## 5 General

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature eg. Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989); *DNA Cloning, Volumes I and*  
10 *ii* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene*  
15 *Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

20 Standard abbreviations for nucleotides and amino acids are used in this specification.

All publications, patents, and patent applications cited herein are incorporated in full by reference. In particular, the contents of UK patent applications 9723516.2, 9724190.5, 9724386.9, 9725158.1, 9726147.3, 9800759.4, and 9819016.8 are incorporated herein.

## Definitions

25 A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term "comprising" means "including" as well as "consisting" eg. a composition "comprising" X may consist exclusively of X or may include something additional to X, such as X+Y.

The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a Neisserial sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature.

An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (eg. 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (eg. see US patent 5,753,235).

#### Expression systems

The Neisserial nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

### i. Mammalian Systems

Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual*, 2nd ed.].

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. (1982b) *Proc. Natl. Acad. Sci.* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally, some enhancers are regulatable and become active only



in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell* 41:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal

viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].

10 The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

15 Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (eg. Hep G2), and a number of other cell lines.

## 20 ii. Baculovirus Systems

The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

30 After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques

are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its own set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* 69:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human  $\alpha$ -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacqz-Verheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter

and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion  
5 can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

10 The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein,  
15 which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15  $\mu$ m in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from  
20 wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers and Smith, *supra*; Miller et al. (1989).

25 Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, *et al.* (1989) *In Vitro Cell. Dev. Biol.* 25:225).

30

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. *See, eg. Summers and Smith supra.*

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as chromatography, eg. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, eg. proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

### iii. Plant Systems

There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsal et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillin, *Gibberellins*: in: *Advanced Plant Physiology*, Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52.

References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for *Agrobacterium* transformations, T DNA sequences for *Agrobacterium*-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Repr.*, 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet*, 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl. Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.



All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

#### iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *Escherichia coli* (*E. coli*) [Raibaud *et al.* (1984) *Annu. Rev. Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature* 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl. Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The g-laotamase (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21].

Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA  
5 polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci.* 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for  
10 the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979)  
15 "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in *Escherichia coli*." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked  
20 with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* or *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the  
25 N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene  
30 [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.*

(1989) *J. Gen. Microbiol.* 135:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (eg. ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology* 7:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghrayeb *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal

element (eg. plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various *Bacillus* strains integrate into the *Bacillus* chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia coli* [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907],

*Streptococcus cremoris* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; *Streptococcus lividans* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], *Streptomyces lividans* [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with  $\text{CaCl}_2$  or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See *eg.* [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, *Bacillus*], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, *Campylobacter*], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of *Escherichia coli* with ColE1-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; *Escherichia*], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 *Lactobacillus*]; [Fiedler *et al.* (1988) *Anal. Biochem* 170:38, *Pseudomonas*]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, *Staphylococcus*], [Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of *Streptococcus lactis* by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Evr. Cong. Biotechnology* 1:412, *Streptococcus*].

#### v. Yeast Expression

Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence

of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1].

In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119; Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *eg.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*eg.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (*eg.* see WO 89/02463.)



Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein *et al.* (1979) *Gene* 8:17-24], pCI/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See eg. Brake *et al.*, *supra*.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the

chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol. Rev.* 51:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol. Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.* (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.* (1990) *Bio/Technology* 8:135], *Pichia guilliermondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555], *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *eg.* [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*];

[Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; Hansenula]; [Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135; Kluyveromyces]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patent  
5 Nos. 4,837,148 and 4,929,555; Pichia]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163 *Saccharomyces*]; [Beach and Nurse (1981) *Nature* 300:706; *Schizosaccharomyces*]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; *Yarrowia*].

### Antibodies

10 As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, humanised  
15 antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying Neisserial proteins.

Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably  
20 a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 µg/injection  
25 is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by *in vitro* immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating  
30 the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is

recovered by centrifugation (eg. 1,000g for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [*Nature* (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as described  
5 above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of  
10 the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (eg. hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then  
15 cultured either *in vitro* (eg. in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly  $^{32}\text{P}$  and  $^{125}\text{I}$ ), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes  
20 are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A,  
25 and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example,  $^{125}\text{I}$  may serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of  
30 this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with  $^{125}\text{I}$ , or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be

readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

### Pharmaceutical Compositions

Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

### Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (*eg.* see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

### Vaccines

Vaccines according to the invention may either be prophylactic (*ie.* to prevent infection) or therapeutic (*ie.* to treat disease after infection).

Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, *etc.* pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, *etc.*; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents

such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required)

5 formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial

10 cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (*eg.*

15 IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, *etc.*), interferons (*eg.* gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), *etc.*; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59™ are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), *etc.*

20

The immunogenic compositions (*eg.* the immunising antigen/immunogen/polypeptide/protein/nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such

25 as water, saline, glycerol, ethanol, *etc.* Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for

30 enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (*eg.* nonhuman primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, *eg.* by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (*eg.* WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [*eg.* Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; see later herein].

#### Gene Delivery Vehicles

Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence *in vivo* can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus,



picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

5 Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses eg. MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

10 Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

15 These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

20 Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (eg. HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

25 Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or

collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654.

Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the

native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

- 5 Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, *Nature* 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317;
- 10 Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805;
- 15 Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240;
- 20 Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245;
- 25 Tonate virus, for example ATCC VR-925; Trinita virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre
- 30 (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid

expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

- 10 Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 15 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the 20 beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional 25 vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active 30 promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA*

91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, *Biochemistry*, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

#### Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for expression of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in eg. WO93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

#### Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

##### A. Polypeptides

One example are polypeptides which include, without limitation: asialoglycoprotein (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

##### B. Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

##### C. Polyalkylenes, Polysaccharides, etc.

Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

##### D. Lipids, and Liposomes

The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the

use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta*. 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleoyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, eg. Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilammellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See eg. Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348; Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.



### E.Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, and E, over time these lipoproteins lose A and acquire C and E apoproteins. VLDL comprises A, B, C, and E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, and E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem* 261:12918; Kane (1980) *Proc Natl Acad Sci USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol. (supra)*; Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30:

443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

#### F. Polycationic Agents

- 5 Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both in vitro, ex vivo, and in vivo applications. Polycationic agents can  
10 be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful  
15 as nucleic acid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the  
20 list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

#### 25 Immunodiagnostic Assays

Neisserial antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-Neisserial antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods. Antibodies to Neisserial proteins within biological samples, including for example, blood or serum

samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the compositions of the invention, in suitable containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, *etc.*) required for the conduct of the assay, as well as suitable set of assay instructions.

#### Nucleic Acid Hybridisation

“Hybridization” refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] Volume 2, chapter 9, pages 9.47 to 9.57.

“Stringency” refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated  $T_m$  of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The

total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to  $10^{-9}$  to  $10^{-8}$  g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of  $10^8$  cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than  $10^8$  cpm/µg, resulting in an exposure time of ~24 hours.

- Several factors can affect the melting temperature ( $T_m$ ) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10} C_i) + 0.4[\%(G + C)] - 0.6(\%\text{formamide}) - 600/n - 1.5(\%\text{mismatch}).$$

where  $C_i$  is the salt concentration (monovalent ions) and  $n$  is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

- In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

- In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology,

and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

#### Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

The nucleic acid probes will hybridize to the Neisserial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native Neisserial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the Neisserial sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Neisserial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a Neisserial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Neisserial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as temperature, salt condition and the like. For example, for diagnostic applications, depending on the

complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

- 5 Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *eg.*  
10 backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*eg.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as peptide nucleic acids may also be used [*eg.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].

- 15 Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [*Meth. Enzymol.* (1987) 155: 335-350]; US patents 4,683,195 and 4,683,202. Two "primer" nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement) to aid with  
20 duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Neisserial sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern  
25 blots. When using the Southern blot method, the labelled probe will hybridize to the Neisserial sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid  
30 support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed

to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1-20 show biochemical data obtained in the Examples, and also sequence analysis, for ORFs 37, 5, 2, 15, 22, 28, 32, 4, 61, 76, 89, 97, 106, 138, 23, 25, 27, 79, 85 and 132. M1 and M2 are molecular weight markers. Arrows indicate the position of the main recombinant product or, in Western blots, the position of the main *N.meningitidis* immunoreactive band. TP indicates *N.meningitidis* total protein extract; OMV indicates *N.meningitidis* outer membrane vesicle preparation. In bactericidal assay results: a diamond (♦) shows preimmune data; a triangle (▲) shows GST control data; a circle (●) shows data with recombinant *N.meningitidis* protein. Computer analyses show a hydrophilicity plot (upper), an antigenic index plot (middle), and an AMPHI analysis (lower). The AMPHI program has been used to predict T-cell epitopes [Gao *et al.* (1989) *J. Immunol.* **143**:3007; Roberts *et al.* (1996) *AIDS Res Hum Retrovir* **12**:593; Quakyi *et al.* (1992) *Scand J Immunol* suppl.11:9) and is available in the Protean package of DNASTAR, Inc. (1228 South Park Street, Madison, Wisconsin 53715 USA).

## EXAMPLES

The examples describe nucleic acid sequences which have been identified in *N.meningitidis*, along with their putative translation products, and also those of *N.gonorrhoeae*. Not all of the nucleic acid sequences are complete *ie.* they encode less than the full-length wild-type protein.

The examples are generally in the following format:

- a nucleotide sequence which has been identified in *N.meningitidis* (strain B)
- the putative translation product of this sequence
- a computer analysis of the translation product based on database comparisons
- corresponding gene and protein sequences identified in *N.meningitidis* (strain A) and in *N.gonorrhoeae*
- a description of the characteristics of the proteins which indicates that they might be suitably antigenic
- results of biochemical analysis (expression, purification, ELISA, FACS *etc.*)

The examples typically include details of sequence identity between species and strains. Proteins that are similar in sequence are generally similar in both structure and function, and the sequence identity often indicates a common evolutionary origin. Comparison with sequences of proteins of known function is widely used as a guide for the assignment of putative protein function to a new sequence and has proved particularly useful in whole-genome analyses.

Sequence comparisons were performed at NCBI (<http://www.ncbi.nlm.nih.gov>) using the algorithms BLAST, BLAST2, BLASTn, BLASTp, tBLASTn, BLASTx, & tBLASTx [eg. see also Altschul *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:2289-3402]. Searches were performed against the following databases: non-redundant GenBank+EMBL+DDBJ+PDB sequences and non-redundant GenBank CDS translations+PDB+SwissProt+SPupdate+PIR sequences.

To compare Meningococcal and Gonococcal sequences, the tBLASTx algorithm was used, as implemented at [http://www.genome.ou.edu/gono\\_blast.html](http://www.genome.ou.edu/gono_blast.html). The FASTA algorithm was also used to compare the ORFs (from GCG Wisconsin Package, version 9.0).

Dots within nucleotide sequences (eg. position 495 in SEQ ID 11) represent nucleotides which have been arbitrarily introduced in order to maintain a reading frame. In the same way, double-underlined nucleotides were removed. Lower case letters (eg. position 496 in SEQ ID 11) represent ambiguities which arose during alignment of independent sequencing reactions (some of the nucleotide sequences in the examples are derived from combining the results of two or more experiments).

Nucleotide sequences were scanned in all six reading frames to predict the presence of hydrophobic domains using an algorithm based on the statistical studies of Esposti *et al.* [Critical evaluation of the hydropathy of membrane proteins (1990) *Eur J Biochem* 190:207-219]. These domains represent potential transmembrane regions or hydrophobic leader sequences.

Open reading frames were predicted from fragmented nucleotide sequences using the program ORFFINDER (NCBI).

Underlined amino acid sequences indicate possible transmembrane domains or leader sequences in the ORFs, as predicted by the PSORT algorithm (<http://www.psорт.nibb.ac.jp>). Functional domains were also predicted using the MOTIFS program (GCG Wisconsin & PROSITE).



Various tests can be used to assess the *in vivo* immunogenicity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *ie.* the protein is an immunogen. This method can also be used to identify immunodominant proteins.

The recombinant protein can also be conveniently used to prepare antibodies *eg.* in a mouse. These can be used for direct confirmation that a protein is located on the cell-surface. Labelled antibody (*eg.* fluorescent labelling for FACS) can be incubated with intact bacteria and the presence of label on the bacterial surface confirms the location of the protein.

10 In particular, the following methods (A) to (S) were used to express, purify and biochemically characterise the proteins of the invention:

#### A) Chromosomal DNA preparation

*N.meningitidis* strain 2996 was grown to exponential phase in 100ml of GC medium, harvested by centrifugation, and resuspended in 5ml buffer (20% Sucrose, 50mM Tris-HCl, 50mM EDTA, pH8).  
15 After 10 minutes incubation on ice, the bacteria were lysed by adding 10ml lysis solution (50mM NaCl, 1% Na-Sarkosyl, 50µg/ml Proteinase K), and the suspension was incubated at 37°C for 2 hours. Two phenol extractions (equilibrated to pH 8) and one  $\text{ChCl}_3$ /isoamylalcohol (24:1) extraction were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes ethanol, and was collected by centrifugation. The pellet was washed once with 70%  
20 ethanol and redissolved in 4ml buffer (10mM Tris-HCl, 1mM EDTA, pH 8). The DNA concentration was measured by reading the OD at 260 nm.

#### B) Oligonucleotide design

Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF, using (a) the meningococcus B sequence when available, or (b) the gonococcus/meningococcus A  
25 sequence, adapted to the codon preference usage of meningococcus as necessary. Any predicted signal peptides were omitted, by deducing the 5'-end amplification primer sequence immediately downstream from the predicted leader sequence.

For most ORFs, the 5' primers included two restriction enzyme recognition sites (*Bam*HI-*Nde*I, *Bam*HI-*Nhe*I, or *Eco*RI-*Nhe*I, depending on the gene's own restriction pattern); the 3' primers included

a *XhoI* restriction site. This procedure was established in order to direct the cloning of each amplification product (corresponding to each ORF) into two different expression systems: pGEX-KG (using either *BamHI-XhoI* or *EcoRI-XhoI*), and pET21b+ (using either *NdeI-XhoI* or *NheI-XhoI*).

5'-end primer tail: CGCGGATCCCATATG (*BamHI-NdeI*)

5 CGCGGATCCGCTAGC (*BamHI-NheI*)

CCGGAATTCTAGCTAGC (*EcoRI-NheI*)

3'-end primer tail: CCCGCTCGAG (*XhoI*)

For ORFs 5, 15, 17, 19, 20, 22, 27, 28, 65 & 89, two different amplifications were performed to clone each ORF in the two expression systems. Two different 5' primers were used for each ORF; the same 3' *XhoI* primer was used as before:

5'-end primer tail: GGAATTCCATATGGCCATGG (*NdeI*)

5'-end primer tail: CGGGATCC (*BamHI*)

ORF 76 was cloned in the pTRC expression vector and expressed as an amino-terminus His-tag fusion. In this particular case, the predicted signal peptide was included in the final product. *NheI*-

15 *BamHI* restriction sites were incorporated using primers:

5'-end primer tail: GATCAGCTAGCCATATG (*NheI*)

3'-end primer tail: CGGGATCC (*BamHI*)

As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridized to the sequence to be amplified. The number of hybridizing nucleotides depended on the melting temperature of the whole primer, and was determined for each primer using the formulae:

$$T_m = 4 (G+C) + 2 (A+T) \quad (\text{tail excluded})$$

$$T_m = 64.9 + 0.41 (\% \text{ GC}) - 600/N \quad (\text{whole primer})$$

The average melting temperature of the selected oligos were 65-70°C for the whole oligo and 50-55°C for the hybridising region alone.

Table I (page 487) shows the forward and reverse primers used for each amplification. In certain cases, it will be noted that the sequence of the primer does not exactly match the sequence in the ORF. When initial amplifications were performed, the complete 5' and/or 3' sequence was not

known for some meningococcal ORFs, although the corresponding sequences had been identified in gonococcus. For amplification, the gonococcal sequences could thus be used as the basis for primer design, altered to take account of codon preference. In particular, the following codons were changed: ATA→ATT; TCG→TCT; CAG→CAA; AAG→AAA; GAG→GAA; CGA→CGC; CGG→CGC; GGG→GGC. Italicised nucleotides in Table I indicate such a change. It will be appreciated that, once the complete sequence has been identified, this approach is generally no longer necessary.

Oligos were synthesized by a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2ml NH<sub>4</sub>OH, and deprotected by 5 hours incubation at 56°C. The oligos were precipitated by addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were then centrifuged and the pellets resuspended in either 100µl or 1ml of water. OD<sub>260</sub> was determined using a Perkin Elmer Lambda Bio spectrophotometer and the concentration was determined and adjusted to 2-10pmol/µl.

### C) Amplification

The standard PCR protocol was as follows: 50-200ng of genomic DNA were used as a template in the presence of 20-40µM of each oligo, 400-800µM dNTPs solution, 1x PCR buffer (including 1.5mM MgCl<sub>2</sub>), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaq, GIBCO Platinum, Pwo DNA polymerase, or Tahara Shuzo Taq polymerase).

In some cases, PCR was optimised by the addition of 10µl DMSO or 50µl 2M betaine.

After a hot start (adding the polymerase during a preliminary 3 minute incubation of the whole mix at 95°C), each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridization temperature the one of the oligos excluding the restriction enzymes tail, followed by 30 cycles performed according to the hybridization temperature of the whole length oligos. The cycles were followed by a final 10 minute extension step at 72°C.

The standard cycles were as follows:

	Denaturation	Hybridisation	Elongation
First 5 cycles	30 seconds 95°C	30 seconds 50-55°C	30-60 seconds 72°C
Last 30 cycles	30 seconds	30 seconds	30-60 seconds

	95°C	65-70°C	72°C
--	------	---------	------

The elongation time varied according to the length of the ORF to be amplified.

The amplifications were performed using either a 9600 or a 2400 Perkin Elmer GeneAmp PCR System. To check the results, 1/10 of the amplification volume was loaded onto a 1-1.5% agarose gel and the size of each amplified fragment compared with a DNA molecular weight marker.

- 5 The amplified DNA was either loaded directly on a 1% agarose gel or first precipitated with ethanol and resuspended in a suitable volume to be loaded on a 1% agarose gel. The DNA fragment corresponding to the right size band was then eluted and purified from gel, using the Qiagen Gel Extraction Kit, following the instructions of the manufacturer. The final volume of the DNA fragment was 30µl or 50µl of either water or 10mM Tris, pH 8.5.

#### 10 **D) Digestion of PCR fragments**

The purified DNA corresponding to the amplified fragment was split into 2 aliquots and double-digested with:

- *NdeI/XhoI* or *NheI/XhoI* for cloning into pET-21b+ and further expression of the protein as a C-terminus His-tag fusion
- 15 – *BamHI/XhoI* or *EcoRI/XhoI* for cloning into pGEX-KG and further expression of the protein as N-terminus GST fusion.
- For ORF 76, *NheI/BamHI* for cloning into pTRC-HisA vector and further expression of the protein as N-terminus His-tag fusion.
- *EcoRI/PstI*, *EcoRI/SalI*, *Sall/PstI* for cloning into pGex-His and further expression of
- 20 the protein as N-terminus His-tag fusion

- Each purified DNA fragment was incubated (37°C for 3 hours to overnight) with 20 units of each restriction enzyme (New England Biolabs ) in a either 30 or 40µl final volume in the presence of the appropriate buffer. The digestion product was then purified using the QIAquick PCR purification kit, following the manufacturer's instructions, and eluted in a final volume of 30 or
- 25 50µl of either water or 10mM Tris-HCl, pH 8.5. The final DNA concentration was determined by 1% agarose gel electrophoresis in the presence of titrated molecular weight marker.

**E) Digestion of the cloning vectors (pET22B, pGEX-KG, pTRC-His A, and pGex-His)**

10µg plasmid was double-digested with 50 units of each restriction enzyme in 200µl reaction volume in the presence of appropriate buffer by overnight incubation at 37°C. After loading the whole digestion on a 1% agarose gel, the band corresponding to the digested vector was purified  
5 from the gel using the Qiagen QIAquick Gel Extraction Kit and the DNA was eluted in 50µl of 10mM Tris-HCl, pH 8.5. The DNA concentration was evaluated by measuring OD<sub>260</sub> of the sample, and adjusted to 50µg/µl. 1µl of plasmid was used for each cloning procedure.

The vector pGEX-His is a modified pGEX-2T vector carrying a region encoding six histidine residues upstream to the thrombin cleavage site and containing the multiple cloning site of the  
10 vector pTRC99 (Pharmacia).

**F) Cloning**

The fragments corresponding to each ORF, previously digested and purified, were ligated in both pET22b and pGEX-KG. In a final volume of 20µl, a molar ratio of 3:1 fragment/vector was ligated using 0.5µl of NEB T4 DNA ligase (400 units/µl), in the presence of the buffer supplied by the manufacturer.  
15 The reaction was incubated at room temperature for 3 hours. In some experiments, ligation was performed using the Boehringer "Rapid Ligation Kit", following the manufacturer's instructions.

In order to introduce the recombinant plasmid in a suitable strain, 100µl *E. coli* DH5 competent cells were incubated with the ligase reaction solution for 40 minutes on ice, then at 37°C for 3 minutes, then, after adding 800µl LB broth, again at 37°C for 20 minutes. The cells were then  
20 centrifuged at maximum speed in an Eppendorf microfuge and resuspended in approximately 200µl of the supernatant. The suspension was then plated on LB ampicillin (100mg/ml).

The screening of the recombinant clones was performed by growing 5 randomly-chosen colonies overnight at 37°C in either 2ml (pGEX or pTC clones) or 5ml (pET clones) LB broth + 100µg/ml ampicillin. The cells were then pelleted and the DNA extracted using the Qiagen QIAprep Spin  
25 Miniprep Kit, following the manufacturer's instructions, to a final volume of 30µl. 5µl of each individual miniprep (approximately 1g) were digested with either *NdeI/XhoI* or *BamHI/XhoI* and the whole digestion loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in parallel with the molecular weight marker (1Kb DNA Ladder, GIBCO). The screening of the positive clones was made on the base of the correct insert size.

For the cloning of ORFs 110, 111, 113, 115, 119, 122, 125 & 130, the double-digested PCR product was ligated into double-digested vector using *EcoRI*-*PstI* cloning sites or, for ORFs 115 & 127, *EcoRI*-*SaII* or, for ORF 122, *SaII*-*PstI*. After cloning, the recombinant plasmids were introduced in the *E.coli* host W3110. Individual clones were grown overnight at 37°C in L-broth with 50µl/ml ampicillin.

### G) Expression

Each ORF cloned into the expression vector was transformed into the strain suitable for expression of the recombinant protein product. 1µl of each construct was used to transform 30µl of *E.coli* BL21 (pGEX vector), *E.coli* TOP 10 (pTRC vector) or *E.coli* BL21-DE3 (pET vector), as described above. In the case of the pGEX-His vector, the same *E.coli* strain (W3110) was used for initial cloning and expression. Single recombinant colonies were inoculated into 2ml LB+Amp (100µg/ml), incubated at 37°C overnight, then diluted 1:30 in 20ml of LB+Amp (100µg/ml) in 100ml flasks, making sure that the OD<sub>600</sub> ranged between 0.1 and 0.15. The flasks were incubated at 30°C into gyratory water bath shakers until OD indicated exponential growth suitable for induction of expression (0.4-0.8 OD for pET and pTRC vectors; 0.8-1 OD for pGEX and pGEX-His vectors). For the pET, pTRC and pGEX-His vectors, the protein expression was induced by addition of 1mM IPTG, whereas in the case of pGEX system the final concentration of IPTG was 0.2mM. After 3 hours incubation at 30°C, the final concentration of the sample was checked by OD. In order to check expression, 1ml of each sample was removed, centrifuged in a microfuge, the pellet resuspended in PBS, and analysed by 12% SDS-PAGE with Coomassie Blue staining. The whole sample was centrifuged at 6000g and the pellet resuspended in PBS for further use.

### H) GST-fusion proteins large-scale purification.

A single colony was grown overnight at 37°C on LB+Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture in a water bath shaker and grown overnight. Bacteria were diluted 1:30 into 600ml of fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD<sub>550</sub> 0.8-1. Protein expression was induced with 0.2mM IPTG followed by three hours incubation. The culture was centrifuged at 8000rpm at 4°C. The supernatant was discarded and the bacterial pellet was resuspended in 7.5ml cold PBS. The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and centrifuged again. The supernatant was collected and mixed with 150µl Glutathione-Sepharose 4B resin (Pharmacia)

(previously washed with PBS) and incubated at room temperature for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml cold PBS for 10 minutes, resuspended in 1ml cold PBS, and loaded on a disposable column. The resin was washed twice with 2ml cold PBS until the flow-through reached OD<sub>280</sub> of 0.02-0.06. The GST-fusion protein was eluted by addition of 700µl cold Glutathione elution buffer (10mM reduced glutathione, 50mM Tris-HCl) and fractions collected until the OD<sub>280</sub> was 0.1. 21µl of each fraction were loaded on a 12% SDS gel using either Biorad SDS-PAGE Molecular weight standard broad range (M1) (200, 116.25, 97.4, 66.2, 45, 31, 21.5, 14.4, 6.5 kDa) or Amersham Rainbow Marker (M2) (220, 66, 46, 30, 21.5, 14.3 kDa) as standards. As the MW of GST is 26kDa, this value must be added to the MW of each GST-fusion protein.

#### **I) His-fusion solubility analysis (ORFs 111-129)**

To analyse the solubility of the His-fusion expression products, pellets of 3ml cultures were resuspended in buffer M1 [500µl PBS pH 7.2]. 25µl lysozyme (10mg/ml) was added and the bacteria were incubated for 15 min at 4°C. The pellets were sonicated for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and then separated again into pellet and supernatant by a centrifugation step. The supernatant was collected and the pellet was resuspended in buffer M2 [8M urea, 0.5M NaCl, 20mM imidazole and 0.1M NaH<sub>2</sub>PO<sub>4</sub>] and incubated for 3 to 4 hours at 4°C. After centrifugation, the supernatant was collected and the pellet was resuspended in buffer M3 [6M guanidinium-HCl, 0.5M NaCl, 20mM imidazole and 0.1M NaH<sub>2</sub>PO<sub>4</sub>] overnight at 4°C. The supernatants from all steps were analysed by SDS-PAGE.

The proteins expressed from ORFs 113, 119 and 120 were found to be soluble in PBS, whereas ORFs 111, 122, 126 and 129 need urea and ORFs 125 and 127 need guanidinium-HCl for their solubilization.

#### **J) His-fusion large-scale purification.**

A single colony was grown overnight at 37°C on a LB + Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture and incubated overnight in a water bath shaker. Bacteria were diluted 1:30 into 600ml fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD<sub>550</sub> 0.6-0.8. Protein expression was induced by addition of 1mM IPTG and the culture further incubated for three hours. The culture was centrifuged at 8000rpm at 4°C, the supernatant was discarded and the bacterial pellet was resuspended in 7.5ml of either (i) cold

buffer A (300mM NaCl, 50mM phosphate buffer, 10mM imidazole, pH 8) for soluble proteins or (ii) buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 8.8) for insoluble proteins.

The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed two times and centrifuged again.

- 5 For insoluble proteins, the supernatant was stored at -20°C, while the pellets were resuspended in 2ml buffer C (6M guanidine hydrochloride, 100mM phosphate buffer, 10mM Tris-HCl, pH 7.5) and treated in a homogenizer for 10 cycles. The product was centrifuged at 13000rpm for 40 minutes.

Supernatants were collected and mixed with 150µl Ni<sup>2+</sup>-resin (Pharmacia) (previously washed with either buffer A or buffer B, as appropriate) and incubated at room temperature with gentle agitation  
10 for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml buffer A or B for 10 minutes, resuspended in 1ml buffer A or B and loaded on a disposable column. The resin was washed at either (i) 4°C with 2ml cold buffer A or (ii) room temperature with 2ml buffer B, until the flow-through reached OD<sub>280</sub> of 0.02-0.06.

The resin was washed with either (i) 2ml cold 20mM imidazole buffer (300mM NaCl, 50mM  
15 phosphate buffer, 20mM imidazole, pH 8) or (ii) buffer D (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 6.3) until the flow-through reached the O.D<sub>280</sub> of 0.02-0.06. The His-fusion protein was eluted by addition of 700µl of either (i) cold elution buffer A (300mM NaCl, 50mM phosphate buffer, 250mM imidazole, pH 8) or (ii) elution buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 4.5) and fractions collected until the O.D<sub>280</sub> was 0.1. 21µl of each  
20 fraction were loaded on a 12% SDS gel.

### K) His-fusion proteins renaturation

10% glycerol was added to the denatured proteins. The proteins were then diluted to 20µg/ml using dialysis buffer I (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, 2M urea, pH 8.8) and dialysed against the same buffer at 4°C for 12-  
25 14 hours. The protein was further dialysed against dialysis buffer II (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was evaluated using the formula:

$$\text{Protein (mg/ml)} = (1.55 \times \text{OD}_{280}) - (0.76 \times \text{OD}_{260})$$



**L) His-fusion large-scale purification (ORFs 111-129)**

500ml of bacterial cultures were induced and the fusion proteins were obtained soluble in buffer M1, M2 or M3 using the procedure described above. The crude extract of the bacteria was loaded onto a Ni-NTA superflow column (Quiagen) equilibrated with buffer M1, M2 or M3 depending on the solubilization buffer of the fusion proteins. Unbound material was eluted by washing the column with the same buffer. The specific protein was eluted with the corresponding buffer containing 500mM imidazole and dialysed against the corresponding buffer without imidazole. After each run the columns were sanitized by washing with at least two column volumes of 0.5 M sodium hydroxide and reequilibrated before the next use.

**10 M) Mice immunisations**

20µg of each purified protein were used to immunise mice intraperitoneally. In the case of ORFs 2, 4, 15, 22, 27, 28, 37, 76, 89 and 97, Balb-C mice were immunised with Al(OH)<sub>3</sub> as adjuvant on days 1, 21 and 42, and immune response was monitored in samples taken on day 56. For ORFs 44, 106 and 132, CD1 mice were immunised using the same protocol. For ORFs 25 and 40, CD1 mice were immunised using Freund's adjuvant, rather than Al(OH)<sub>3</sub>, and the same immunisation protocol was used, except that the immune response was measured on day 42, rather than 56. Similarly, for ORFs 23, 32, 38 and 79, CD1 mice were immunised with Freund's adjuvant, but the immune response was measured on day 49.

**N) ELISA assay (sera analysis)**

20 The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 7ml of Mueller-Hinton Broth (Difco) containing 0.25% Glucose. Bacterial growth was monitored every 30 minutes by following OD<sub>620</sub>. The bacteria were let to grow until the OD reached the value of 0.3-0.4. The culture was centrifuged for 10 minutes at 10000rpm. The supernatant was discarded and bacteria were washed once with PBS, resuspended in PBS containing 0.025% formaldehyde, and incubated for 2 hours at room temperature and then overnight at 4°C with stirring. 100µl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing buffer (0.1% Tween-20 in PBS). 200µl of saturation buffer (2.7% Polyvinylpyrrolidone 10 in water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed

three times with PBT. 200µl of diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN<sub>3</sub> in PBS) were added to each well and the plates incubated for 90 minutes at 37°C. Wells were washed three times with PBT. 100µl of HRP-conjugated rabbit anti-mouse (Dako) serum diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C. Wells were washed three times with PBT buffer. 100µl of substrate buffer for HRP (25ml of citrate buffer pH5, 10mg of O-phenildiamine and 10µl of H<sub>2</sub>O) were added to each well and the plates were left at room temperature for 20 minutes. 100µl H<sub>2</sub>SO<sub>4</sub> was added to each well and OD<sub>490</sub> was followed. The ELISA was considered positive when OD<sub>490</sub> was 2.5 times the respective pre-immune sera.

#### 10 O) FACScan bacteria Binding Assay procedure.

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD<sub>620</sub>. The bacteria were let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and the pellet was resuspended in blocking buffer (1% BSA, 0.4% NaN<sub>3</sub>) and centrifuged for 5 minutes at 4000rpm. Cells were resuspended in blocking buffer to reach OD<sub>620</sub> of 0.07. 100µl bacterial cells were added to each well of a Costar 96 well plate. 100µl of diluted (1:200) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000rpm, the supernatant aspirated and cells washed by addition of 200µl/well of blocking buffer in each well. 100µl of R-Phicoerytrin conjugated F(ab)<sub>2</sub> goat anti-mouse, diluted 1:100, was added to each well and plates incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and washed by addition of 200µl/well of blocking buffer. The supernatant was aspirated and cells resuspended in 200µl/well of PBS, 0.25% formaldehyde. Samples were transferred to FACScan tubes and read. The condition for FACScan setting were: FL1 on, FL2 and FL3 off; FSC-H threshold:92; FSC PMT Voltage: E 02; SSC PMT: 474; Amp. Gains 7.1; FL-2 PMT: 539; compensation values: 0.

**P) OMV preparations**

Bacteria were grown overnight on 5 GC plates, harvested with a loop and resuspended in 10 ml 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes and the bacteria disrupted by sonication for 10 minutes on ice (50% duty cycle, 50% output). Unbroken cells were removed by centrifugation at 5000g for 10 minutes and the total cell envelope fraction recovered by centrifugation at 50000g at 4°C for 75 minutes. To extract cytoplasmic membrane proteins from the crude outer membranes, the whole fraction was resuspended in 2% sarkosyl (Sigma) and incubated at room temperature for 20 minutes. The suspension was centrifuged at 10000g for 10 minutes to remove aggregates, and the supernatant further ultracentrifuged at 50000g for 75 minutes to pellet the outer membranes. The outer membranes were resuspended in 10mM Tris-HCl, pH8 and the protein concentration measured by the Bio-Rad Protein assay, using BSA as a standard.

**Q) Whole Extracts preparation**

Bacteria were grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes.

**15 R) Western blotting**

Purified proteins (500ng/lane), outer membrane vesicles (5µg) and total cell extracts (25µg) derived from MenB strain 2996 were loaded on 15% SDS-PAGE and transferred to a nitrocellulose membrane. The transfer was performed for 2 hours at 150mA at 4°C, in transferring buffer (0.3 % Tris base, 1.44 % glycine, 20% methanol). The membrane was saturated by overnight incubation at 4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated for 2 hours at 37°C with mice sera diluted 1:200 in washing buffer. The membrane was washed twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labelled anti-mouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

**S) Bactericidal assay**

MC58 strain was grown overnight at 37°C on chocolate agar plates. 5-7 colonies were collected and used to inoculate 7ml Mueller-Hinton broth. The suspension was incubated at 37°C on a nutator and let to grow until OD<sub>620</sub> was 0.5-0.8. The culture was aliquoted into sterile 1.5ml Eppendorf

tubes and centrifuged for 20 minutes at maximum speed in a microfuge. The pellet was washed once in Gey's buffer (Gibco) and resuspended in the same buffer to an OD<sub>620</sub> of 0.5, diluted 1:20000 in Gey's buffer and stored at 25°C.

50µl of Gey's buffer/1% BSA was added to each well of a 96-well tissue culture plate. 25µl of diluted mice sera (1:100 in Gey's buffer/0.2% BSA) were added to each well and the plate incubated at 4°C. 25µl of the previously described bacterial suspension were added to each well. 25µl of either heat-inactivated (56°C waterbath for 30 minutes) or normal baby rabbit complement were added to each well. Immediately after the addition of the baby rabbit complement, 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 0). The 96-well plate was incubated for 1 hour at 37°C with rotation and then 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 1). After overnight incubation the colonies corresponding to time 0 and time 1 hour were counted.

Table II (page 493) gives a summary of the cloning, expression and purification results.

### Example 1

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 1>:

```

1  ATGAAACAGA CAGTCAA.AT GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51  GAACCGACCG GTGTGGNCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
101 A.GCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
151 TAT.TACAAA GGACGCGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
201 GTATCGGCAG CCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
251 GCTGGATGTA TGCCAACGGG CGCGC.GTGC GCCAAGATGA TACCGAAGCG
301 GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
351 CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
401 TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
451 GCCCAAACA ATTTGGGCGT GATGTATGCC GAAAGANCGC GCGTGCGCCA
501 AGACCG...
```

This corresponds to the amino acid sequence <SEQ ID 2; ORF37>:

```

1  MKQTVXMLAA ALIALGLNRP VWXDDVSDFR ENLXAAQGN AAAQYNLGAM
51  YXQRTVRVRD DAEAVRWYRQ PAEQGLAQAO YNLGWMYANG RXVRQDDTEA
101 VRWYRQAAAO GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAA
151 AQNNLGVMYA ERXRVRQD...
```

Further work revealed the complete nucleotide sequence <SEQ ID 3>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51  GAACCGAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
101 AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
151 TATTACAAAG GACGCGGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
201 GTATCGGCAG GCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
251 GCTGGATGTA TGCCAACGGG CGCGGCGTGC GCCAAGATGA TACCGAAGCG
301 GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
351 CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
401 TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
451 GCCCAAACA ATTTGGGCGT GATGTATGCC GAAAGACGCG GCGTGCGCCA
501 AGACCGCGCC CTTGCACAAG AATGGTTTGG CAAGGCTTGT CAAAACGGAG
551 ACCAAGACGG CTGCGACAAT GACCAACGCC TGAAGGCGGG TTATTGA
```

This corresponds to the amino acid sequence <SEQ ID 4; ORF37-1>:

```

1  MKQTVKWLAA ALIALGLNRA VWADDVSDFR ENLQAAQGN AAAQYNLGAM
51 YYKGRGVRRD DAEAVRWYRQ AAEQGLAQAG YNLGWMYANG RGVRQDDTEA
101 VRWYRQAAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAQ
5  151 AQNNLGVMYA ERRGVRQDRA LAQEWFGKAC QNGDQDGCND DQRLKAGY*

```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 5>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51 GAACCAAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAAACCTGC
101 AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AAAACAATTT GGGCGTGATG
10  151 TATGCCGAAA GACGCGGCGT GCGCCAAGAC CGCGCCCTTG CACAAGAATG
201 GCTTGGCAAG GCTTGTCAAA ACGGATACCA AGACAGCTGC GACAATGACC
251 AACGCCTGAA AGCGGGTTAT TGA

```

This encodes a protein having amino acid sequence <SEQ ID 6; ORF37a>:

```

1  MKQTVKWLAA ALIALGLNQA VWADDVSDFR ENLQAAQGN AAAQNNLGVM
15  51 YAERRGVRRD RALAQEWLK ACQNGYQDSC DNDQRLKAGY *

```

The originally-identified partial strain B sequence (ORF37) shows 68.0% identity over a 75aa overlap with ORF37a:

```

20  orf37.pep      10      20      30      40      50      60
      MKQTVXMLAAALIALGLNRPVWXDDVSDFRENLXAAQGNAAAQYNLGAMYXQRTVRVRD
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf37a  MKQTVKWLAAALIALGLNQAVWADDVSDFRENLQAAQGNAAAQNNLGVMYAERRGVRRD
      10      20      30      40      50      60

25  orf37.pep      70      80      90      100     110     120
      DAEAVRWYRQPAEQGLAQAYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG
      | | : | : : |
orf37a  RALAQEWLKACQNGYQDSCDNDQRLKAGYX
      70      80      90

```

30 Further work identified the corresponding gene in *N.gonorrhoeae* <SEQ ID 7>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51 GAACCAAGCG GTGTGGGCGG GTGACGTATC GGATTTTCGG GAAAACCTGC
101 AGgcggaGA ACaggGAAAT GCAGCAGCCC AATTCAATTT GGGCGTGATG
35  151 TATGAAATG GACAAGGAGT TCGTCAAGAT TATGTACAGG CAGTGCAGTG
201 GTATCGCAAG GCTTCAGAAC AAGGGGATGC CCAAGCCCAA TACAATTTGG
251 GCTTGATGTA TTACGATGGA CGCGGCGTGC GCCAAGACCT TGCCTCGCT
301 CAACAATGGC TTGGCAAGGC TTGTCAAAAC GGAGACCAA ACAGCTGCCA
351 CAATGACCAA CGCCTGAAGG CGGGTTATTA A

```

This encodes a protein having amino acid sequence <SEQ ID 8; ORF37ng>:

```

40  1  MKQTVKWLAA ALIALGLNQA VWAGDVSDFR ENLQAAEQGN AAAQFNLGVM
      51 YENGQGVRRD YVQAVQWYRK ASEQGDAAQ YNLGLMYDGR RGVRQDLALA
      101 QQWLKGACQN GDQNSCDNDQ RLKAGY*

```

The originally-identified partial strain B sequence (ORF37) shows 64.9% identity over a 111aa overlap with ORF37ng:

```

45  orf37.pep      10      20      30      40      50      60
      MKQTVXMLAAALIALGLNRPVWXDDVSDFRENLXAAQGNAAAQYNLGAMYXQRTVRVRD
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf37ng  MKQTVKWLAAALIALGLNQAVWAGDVSDFRENLQAAEQGNAAAQFNLGVMYENGQGVRRD
      10      20      30      40      50      60

50  orf37.pep      70      80      90      100     110     120
      DAEAVRWYRQPAEQGLAQAYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG
      : || : || : || : || : || : || : || : || : || : || : ||
orf37ng  YVQAVQWYRKASEQGDAAQAYNLGLMYDGRGVRQDLALAAQQWLKGACQNGDQNSCDNDQ
      70      80      90      100     110     120

orf37.pep      VIYAEGRGVRQDDVEAVRWFRQAAAQGVAAQNNLGVMYAERXVRQD 168
55  orf37ng      RLKAGY 126

```

The complete strain B sequence (ORF37-1) and ORF37ng show 51.5% identity in 198 aa overlap:

		10	20	30	40	50	60
5	orf37-1.pep	MKQTVKWLAAALIALGLNRAVWADDVSDFRENLQAAAQGNAAQYNLGAMYYKGRGVR					
	orf37ng	MKQTVKWLAAALIALGLNQAVWAGDVSDFRENLQAAEQGNAAQYNLGVMYENGQGV					
		10	20	30	40	50	60
10	orf37-1.pep	DAEAVRWYRQAAEQGLAQYNLGWMYANGRGVRQDDTEAVRWYRQAAQGVVQAQYNLG					
	orf37ng	YVQAVQWYRKASEQGDAAQYNLGLMYDGRGVRQD-----					
		70	80	90			
15	orf37-1.pep	VIYAEGRGVRQDDVEAVRWFRQAAAQGVQAQNNLGVMYAERRGVRQDRALAQEWFGKAC					
	orf37ng	-----LALAQQWLKAC					
						100	
20		190	199				
	orf37-1.pep	QNGDQDQDNDQRLKAGYX					
	orf37ng	QNGDQNSCDNDQRLKAGYX					
		110	120				

- 25 Computer analysis of these amino acid sequences indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF37-1 (11kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 1A shows the results of affinity purification of the GST-fusion protein, and Figure 1B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), FACS analysis (Figure 1C), and a bactericidal assay (Figure 1D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is a useful immunogen.

- 35 Figure 1E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF37-1.

## Example 2

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 9>:

	TTCGGCGA	CATCGGCGGT	TTGAAGGTCA	ATGCCCCCGT	CAAATCCGCA
40	GGCGTATTGG	TCGGGCGCGT	CGGCGCTATC	GGACTTGACC	CGAAATCCTA
	TCAGGCGAGG	GTGCGCCTCG	ATTTGGACGG	CAAGTATCAG	TTCAGCAGCG
	ACGTTTCCGC	GCAAATCCTG	ACTTCsGGAC	TTTTGGGCGA	GCAGTACATC
	GGGCTGCAGC	AGGGCGGCGA	CACGGAAAAC	CTTGCTGCCG	GCGACACCAT
	CTCCGTAACC	AGTTCTGCAA	TGGTCTTGGA	AAACCTTATC	GGCAAATTCA
45	TGACGAGTTT	TGCCGAGAAA	AATGCCGACG	GCGGCAATGC	GGAAAAAGCC
	GCCGAATAA				

This corresponds to the amino acid sequence <SEQ ID 10>:

1 FGDIGGLKVN APVKSAGVLV GRVGAIGLDP KSYQARVRLD LDGKYQFSSD  
51 VSAQILTSGL LGEQYIGLQQ GGDENLAAG DTISVTSSAM VLENLIGKFM

101 TSFAEKNADG GNAEKAAE\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a hypothetical *H.influenzae* protein (ybrd.haein; accession number p45029)

SEQ ID 9 and ybrd.haein show 48.4% aa identity in 122 aa overlap:

```

5      20      30      40      50      60      70
yrbd.h LGIGALVFLGLRVANVQGF AETKSYTVTATFDNIGGLKVRAPLKIGGVVIGRVSAILTDE
N.m      FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
10      80      90      100     110     120     130
yrbd.h KSYLPKVSIAINQEYNEIPENSSLSIKTSGLLGEQYIALTMGFDDGDTAMLKNGSQIQDT
N.m      KSYQARVRDLDDGKY-QFSSDVSAQILTSGLLGEQYIGLQQG---GDTENLAAGDTISVT
15      40      50      60      70      80
yrbd.h TSAMVLEDLIGQFL--YGSKKSDGNEKSESTEQ
N.m      SSAMVLENLIGKFMTSFAEKNADGGNAEKAAEX
20      90      100     110     120

```

Homology with a predicted ORF from *N.gonorrhoeae*

SEQ ID 9 shows 99.2% identity over a 118aa overlap with a predicted ORF from *N. gonorrhoeae*:

```

25      20      30      40      50      60      70
yrbd      GAAAVAFLAFRVAGGAFFGSDKTYAVYADFGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
N.m      FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
30      80      90      100     110     120     130
yrbd      KSYQARVRDLDDGKYQFSSDVSAQILTSGLLGEQYIGLQQGGDTENLAAGDTISVTSSAM
N.m      KSYQARVRDLDDGKYQFSSDVSAQILTSGLLGEQYIGLQQGGDTENLAAGDTISVTSSAM
35      40      50      60      70      80      90
yrbd      VLENLIGKFMTSFAEKNADGGNAEKAAEX
N.m      VLENLIGKFMTSFAEKNADGGNAEKAAEX
40      100     110     120

```

The complete yrbd *H.influenzae* sequence has a leader sequence and it is expected that the full-length homologous *N.meningitidis* protein will also have one. This suggests that it is either a membrane protein, a secreted protein, or a surface protein and that the protein, or one of its epitopes, could be a useful antigen for vaccines or diagnostics.

### Example 3

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 11>:

```

50      1  ..ATTTTGATAT ACCTCATCCG CAAGAATCTA GGTTCCGCCG TCTTCTTCTT
      51  TCAGGAACGC CCCGGAAGG ACGGAAAACC TTTTAAATG GTCAAATTCC
      101 GTCCATGCG CGACGGCTTG TATTCAGACG GCATTCCGCT GCCCGACGGA
      151 GAACGCCTGA CACCGTTCGG CAAAAAATG CGTGCCGC CA GTwTGGACGA
      201 ACTGCCTGAA TTATGGAATA TCTTAAAAGG CGAGATGAGC CTGGTCGGCC
      251 CCCGCCCGCT GCTGATGCAA TATCTGCCGC TGTACGACAA CTTCCAAAAC
      301 CGCCGCCACG AAATGAAACC CGGCATTACC GGCTGGGCGC AGGTCAACGG

```

351 GCGCAACGCg CTTTCGTGGG ACGAAAAATT CGCCTGCGAT GTTTGGTATA  
 401 TCGACCACTT CAGCCTGTGC CTCGACATCA AAATCCTACT GCTGACGGTT  
 451 AAAAAAGTAT TAATCAAGGA AGGGATTTCC GCACAGGGCG AACAA.aCCAT  
 501 GCCCCCTTTC ACAGGAAAAC GCAAACCTCGC CGTCGTGCGT GCGGGCGGAC  
 5 551 ACGGAAAAGT CGTTGCCGCG CTTGCCGCGG CACTCGGCCG GTACAGGGAA  
 601 ATCGTTTTC TGGACGACCG CGCACAAGGC AGCGTCAACG GCTTTTCCGT  
 651 CATCGGCACG ACGCTGCTGC TTGAAAACAG TTTATCGCCC GAACAATACG  
 701 ACGTCGCCGT CGCCGTGCGC AACAAACGCA TCCGCCGCCA AATCGCCGAA  
 751 AAAGCCGCCG CGCTCGGCTT CGCCCTGCCC GTACTGGTTC ATCCGGACGC  
 10 801 GACCGTCTCG CTTTCTGCAA CAGTCGGACA AGGCAGCGTC GTTATGGCGA  
 851 AAGCGGTCG..

This corresponds to the amino acid sequence <SEQ ID 12; ORF3>:

1 ..ILIYLIRKLNLS GSPVFFFQER PGKDGKPFKM VKFRSMRDGL YSDGIPLPDG  
 51 ERLTPFGKKL RAASXDELPE LWNILKGEMS LVGPRPLLMQ YLPLYDNFQN  
 15 101 RRHEMKPGIT GWAQVNGRNA LSWDEKFACD VWYIDHFSLC LDIKILLTV  
 151 KKVLIKEGIS AQEXTMPFF TGRKRLAVVG AGGHGKVVD LAAALGRYRE  
 201 IVFLDDRAQG SVNGFSVIGT TLLLENSLSP EQYDVAVAVG NNRIRRQIAE  
 251 KAAALGFALP VLVHPDATVS PSATVGQGSV VMAKAV..

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 13>:

20 1 ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG  
 51 ACTGATTTTC CTCTCGCCAG TATTTTGTAT TTTGATATAC CTCATCCGCA  
 101 AGAATCTAGG TTCGCCCGTC TTCTTCTTTC AGGAACGCCC CGGAAAGGAC  
 151 GGAAAACCTT TTAATAATGGT CAAATTCGGT TCCATGCGCG ACGCGCTTGA  
 201 TTCAGACGGC ATTCCGCTGC CCGACGGAGA ACGCCTGACA CCGTTCGGCA  
 25 251 AAAAAGTGGC TGCCGCCAGT TTGGACGAAC TGCCCTGAATT ATGGAATATC  
 301 TTAAAAGGCG AGATGAGCCT GGTGCGCCCC CGCCCGCTGC TGATGCAATA  
 351 TCTGCCGCTG TACGACAACT TCCAAAACCG CCGCCACGAA ATGAAACCCG  
 401 GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC  
 451 GAAAAATTCG CCTGCGATGT TTGGTATATC GACCACTTCA GCCTGTGCCT  
 30 501 CGACATCAAA ATCCTACTGC TGACGGTTAA AAAAGTATTA ATCAAGGAAG  
 551 GGATTTCCGC ACAGGGCGAA GCCACCATGC CCCCTTTCAC AGGAAAACGC  
 601 AAATCGCCG TCGTCCGTGC GGGCGGACAC GGAAGATCG TTGCCGACCT  
 651 TGCCGCCGCA CTCGGCCGCT ACAGGGAAT CGTTTCTCTG GACGACCGCG  
 701 CACAAGGCAG CGTCAACGGC TTTTCCGTCA TCGGCACGAC GCTGCTGCTT  
 35 751 GAAAACAGTT TATCGCCCGA ACAATACGAC GTCGCCGTGC CCGTCGGCAA  
 801 CAACCGCATC CGCCGCCAAA TCGCCGAAAA AGCCGCCGCG CTCGGCTTCG  
 851 CCCTGCCCGT TCTGGTTCAT CCGGACGCGA CCGTCTCGCC TTCTGCAACA  
 901 GTCGGACAAG GCAGCGTCGT TATGGCGAAA GCCGTGCTAC AGGCAGGCAG  
 951 CGTATTGAAA GACGGCGTGA TTGTGAACAC TGCCGCCACC GTCGATCAG  
 40 1001 ACTGCCTGCT TAACGCTTTC GTCCACATCA GCCCAGGCGC GCACCTGTCTG  
 1051 GGCAACACGC ATATCGGCGA AGAAAGCTGG ATAGGCACGG GCGCGTGCAG  
 1101 CCGCCAGCAG ATCCGTATCG GCAGCCGCGC AACCATTGGA GCGGGCGCAG  
 1151 TCGTCGTACG CGACGTTTCA GACGGCATGA CCGTCGCGGG CAATCCGGCA  
 1201 AAGCCGCTGC CGCGCAAAAA CCCCAGAGACC TCGACAGCAT AA

45 This corresponds to the amino acid sequence <SEQ ID 14; ORF3-1>:

1 MSKFFKRLFD IVASASGLIF LSPVFLILIY LIRKNLGSPV FFFQERPGKD  
 51 GKPFKMVKFR SMRDALDSG IPLPDGERLT PFGKKLRAAS LDELPELWNI  
 101 LKGEMSLVGP RPLLMQYLPL YDNFQNRHE MKPGITGWAQ VNGRNALSWD  
 50 151 EKFACDVWYI DHFSLCLDIK ILLLTVKKVL IKEGISAQGE ATMPFTGKR  
 201 KLAVVGAGGH GKVVADLAAA LGRYREIVFL DDRAQGSVNG FSVIGTLLL  
 251 ENSLSPEQYD VAVAVGNRRI RRQIAEKAAA LGFALPVLVH PDATVSPSAT  
 301 VGQGSVVMK AVVQAGSVLK DGIVNTAAT VDHDCLLNAF VHISPGAHL  
 351 GNTHIGESW IGTGACSRQQ IRIGSRATIG AGAVVVRDVS DGMTVAGNPA  
 401 KPLPRKNPET STA\*

55 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF3 shows 93.0% identity over a 286aa overlap with an ORF (ORF3a) from strain A of *N. meningitidis*:



-65-

orf3.pep  
 orf3a  
 5  
 orf3.pep  
 orf3a  
 10  
 orf3.pep  
 orf3a  
 15  
 orf3.pep  
 orf3a  
 20  
 orf3.pep  
 orf3a  
 25  
 orf3.pep  
 orf3a  
 30  
 orf3.pep  
 orf3a

MSKFFKRLFDIVASASGLIFLSPVFLILIIYLIRKNLGSPVFFFQERPGKD GPKPFKMVKFR  
 10 20 30 40 50 60

SMRDGLYSDGIPLPDGERLTPFGKKLRAASXDELPELWNILKGEMSLVGPRLMQYLPL  
 40 50 60 70 80 90

SMHDALDSGDGILLPDGERLTPFGKKLRAASLDELPELWNVLKGDMSLVGPRLMQYLPL  
 70 80 90 100 110 120

YDNFQNRHRHEMKPGITGWAQVNGRNALSWDEKFCADVWYIDHFSCLCDIKILLTVKKVL  
 100 110 120 130 140 150

YDNFQNRHRHEMKPGITGWAQVNGRNALSWDERFACDIWYIDHFSCLCDIKILLTVKKVL  
 130 140 150 160 170 180

IKEGISAQGEXTMPPTGKRKLAVVGAGGHGKV VADLAAALGRYREIVFLDDRQGSVNG  
 160 170 180 190 200 210

IKEGISAQGEATMPPTGKRKLAVVGAGGHGKVVAELAAALGTYGEIVFLDDRQGSVNG  
 190 200 210 220 230 240

FSVIGTTLLENLSLSPEQYDVAVAVGNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT  
 220 230 240 250 260 270

FPVIGTTLLENLSLSPEQFDIAVAVGNRIRRQIAEKAAALGFALPVLVHPDSTVSPSAT  
 250 260 270 280 290 300

VGQGSVMAKAV  
 280

VGQGGVMAKAVVQADSVLKDGIVNTAATVDHDCLLDAFVHISPGAHLSGNTRIGEEESW  
 310 320 330 340 350 360

35 The complete length ORF3a nucleotide sequence <SEQ ID 15> is:

1 ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG  
 51 ACTGATTTTC CTCTCGCCAG TATTTTGTAT TTTGATATAC CTCATCCGCA  
 101 AGAATCTGGG TTCGCCGCTC TTCTTCTTTC AGGAACGCCC CGGAAAGGAC  
 151 GGAAAACCTT TTAAAATGGT CAAATTCCGT TCCATGCACG ACGCGCTTGA  
 40 201 TTCAGACGGC ATTCTGCTGC CCGACGGAGA ACGCCTGACA CCGTTCGGCA  
 251 AAAAAGTGGC TGCCGCCAGT TTGGACGAAC TGCCCGAACT GTGGAACGTC  
 301 CTCAAAGGCG ACATGAGCCT GGTCGGCCCC CGCCCGCTGC TGATGCAATA  
 351 TCTGCCGCTG TACGACAACT TCCAAAACCG CCGCCACGAA ATGAAACCGG  
 401 GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC  
 45 451 GAACGCTTCG CATGCGACAT CTGGTATATC GACCACTTCA GCCTGTGCCT  
 501 CGACATCAAA ATCCTACTGC TGACGGTTAA AAAAGTATTA ATCAAAGAAG  
 551 GGATTTCCCG ACAGGGCGAA GCCACCATGC CCCCTTTCAC AGGAAAACGC  
 601 AAACCTGCGG TCGTCGGTGC GGGCGGACAC GGCAAAGTCG TTGCCGAGCT  
 651 TGCCGCCGCA CTCGGCACAT ACGGCGAAAT CGTTTTTCTG GACGACCGCG  
 50 701 TCCAAGGCAG CGTCAACGGC TTCCCCGTCA TCGGCACGAC GCTGCTGCTT  
 751 GAAAACAGTT TATCGCCCGA ACAATTGAC ATCGCCGTCG CCGTCGGCAA  
 801 CAACCGCATC CGCCGCCAAA TCGCCGAAAA AGCCGCGCGC CTCGGCTTCG  
 851 CCCTGCCCCG CTGATTTCAT CCGGACTCGA CCGTCTCGCC TTCTGCAACA  
 901 GTCGGACAAG GCGGCGTCGT TATGGCGAAA GCCGTCGTAC AGGCTGACAG  
 55 951 CGTATTGAAA GACGGCGTAA TTGTGAACAC TGCCGCCACC GTCGATCAGC  
 1001 ATTGCCTGCT TGATGCTTTC GTCCACATCA GCCCGGGCGC GCACCTGTGC  
 1051 GGCAACACGC GTATCGGCGA AGAAAGCTGG ATAGGCACAG GCGCGTGCAG  
 1101 CCGCCAGCAG ATCCGTATCG GCAGCCGCGC AACCATTGGA GCGGCGCAG  
 1151 TCGTCGTGCG CGACGTTTCA GACGGCATGA CCGTCGCGGG CAACCCGGCA  
 60 1201 AAACCATTGG CAGGCAAAAA TACCGAGACC CTGCGGTCGT AA

This is predicted to encode a protein having amino acid sequence <SEQ ID 16>:

1 MSKFFKRLFD IVASASGLIF LSPVFLILII LIRKNLGSPV FFFQERPGKD  
 51 GKPFKMVKFR SMHDALDSDG ILLPDGERLT PFGKKLRAAS LDELPELWNV  
 101 LKGDMSLVGP RPLLMQYLPL YDNFQNRHE MKPGITGWAQ VNGRNALSWD  
 65 151 ERFACDIWYI DHFSLCLDIK ILLLTVKKVL IKEGISAQGE ATMPPTFGKR  
 201 KLAVVGAGGH GKVVAELAAA LGTYGEIVFL DDRVQGSVNG FPFVIGTLLL  
 251 ENSLSPEQFD IAVAVGNRRI RRQIAEKAAA LGFALPVLH PDSTVSPSAT

301 VGQGGVMAK AVVQADSVLK DGIVNTAAT VDHDCLLDAF VHISPGAHLS  
 351 GNTRIGEESW IGTGACSRQQ IRIGSRATIG AGAVVVRDVS DGMTVAGNPA  
 401 KPLAGKNTET LRS\*

Two transmembrane domains are underlined.

5 ORF3-1 shows 94.6% identity in 410 aa overlap with ORF3a:

		10	20	30	40	50	60
	orf3a.pep	MSKFFKRLFDIVASASGLIFLSPVFLILYLRKKNLGSPVFFFQERPGKDGP	KPFKMKVKFR				
	orf3-1	MSKFFKRLFDIVASASGLIFLSPVFLILYLRKKNLGSPVFFFQERPGKDGP	KPFKMKVKFR				
10		10	20	30	40	50	60
	orf3a.pep	SMHDALDSGILLPDGERLTPFGKKLRAASLDELPELWNVLKGDMSLVGPRPLLMQYLPL					
	orf3-1	SMRDALDSGILPLDGERLTPFGKKLRAASLDELPELWNILKGEMSLVGPRPLLMQYLPL					
15		70	80	90	100	110	120
	orf3a.pep	YDNFQNRHHEMKPGITGWAQVNGRNALSWDERFACDIWYIDHFSCLCDIKILLT	TVKKVL				
	orf3-1	YDNFQNRHHEMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSCLCDIKILLT	TVKKVL				
20		130	140	150	160	170	180
	orf3a.pep	IKEGISAQGEATMPFFTGRKRLAVVGAGGHGKVVAAALGTYGEIVFLDDR	VQGSVNG				
	orf3-1	IKEGISAQGEATMPFFTGRKRLAVVGAGGHGKVVADLAAALGRYREIVFLDDRA	QGSVNG				
25		190	200	210	220	230	240
	orf3a.pep	FPVIGTTLLENLSLSPQFDIAVAVGNNRIRRQIAEKAAALGFALPVLHPDST	VSPSAT				
	orf3-1	FSVIGTTLLENLSLSPQYDVAVAVGNNRIRRQIAEKAAALGFALPVLHPDAT	VSPSAT				
30		250	260	270	280	290	300
	orf3a.pep	VGQGGVMAKAVVQADSVLKDGIVNTAATVDHDCLLDAFVHISPGAHLSGNTRIGEESW					
	orf3-1	VGQGSVMAKAVVQAGSVLKDGIVNTAATVDHDCLLNAFVHISPGAHLSGNTHIGEESW					
35		310	320	330	340	350	360
	orf3a.pep	IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLAGKNTETLRSX					
	orf3-1	IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLPRKNPETSTAX					
40		370	380	390	400	410	
	orf3a.pep	W+++KF DVWY+D++S LD EGI T FTG					
	orf3-1	WEKKFELDVWYVDNWSFFLDLKLCLTVRKVLVSEGIQQTNHVTAERFTG					

Homology with hypothetical protein encoded by yvfc gene (accession Z71928) of *B. subtilis*

ORF3 and YVFC proteins show 55% aa identity in 170 aa overlap (BLASTp):

50	ORF3	3	IYLIRKKNLGSPVFFFQERPGKDGP	KPFKMKVKFRSMRDGLYSDGIPLDGERLTPFGKKLRA	62
			I ++R +GSPVFF Q RPG GKPF + KFR+M D S G LPD RLT G+ +R		
	yvfc	27	IADVRLKIGSPVFFKQVRPGLHGKPF	TLYKFRTMTDERDSKGNLLPDEVRLTKTGRLIRK	86
55	ORF3	63	ASXDELPELWNILKGEMSLVGPRPLLMQYLPLYDNFQNRHHEMKPGITGWAQVNGRNALS	122	
			S DELP+L N+LKG++SLVGPRPLLM YLPLY Q RRHE+KPGITGWAQ+NGRNA+S		
	yvfc	87	LSIDELPQLLNVLKGDLSLVGPRPLLMQYLPLYTEKQARRHEVKPGITGWAQINGRNAIS	146	
60	ORF3	123	WDEKFACDVWYIDHFSCLDXXXXXXXEGISAQGEXTMPFFTG	172	
			W+++KF DVWY+D++S LD EGI T FTG		
	yvfc	147	WEKKFELDVWYVDNWSFFLDLKLCLTVRKVLVSEGIQQTNHVTAERFTG	196	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF3 shows 86.3% identity over a 286aa overlap with a predicted ORF (ORF3.ng) from *N.gonorrhoeae*:

5	orf3	ILIIYLIRKNLGSPVFFQERPGKDGPVKMKFR	34
	orf3ng	MSKAVKRLFDIIASASGLIVLSPVFLVLIYLIRKNGSPVFFIRERPGKDGPVKMKFR	60
	orf3	SMRDGLYSDGIPDPGERLTPFGKKLRAASXDELPELWNILKGEMSLVGPRPLLMQYLPL	94
10	orf3ng	SMRDALDSGIPDPDSERLTDGFKKLRLATSLDELPELWNVLKGEMSLVGPRPLLMQYLPL	120
	orf3	YDNFQNRHRHEMKPGITGWAQVNGRNALSWDEKFCADVWYIDHFSCLCDIKILLTVKKVL	154
15	orf3ng	YNKFQNRHRHEMKPGITGWAQVNGRNALSWDEKFCSDVWYTDNFSFWLDMKILFLTVKKVL	180
	orf3	IKEGISAQGEATMPPFTGKRKLAVVGAGGHGKVADLAAALGRYREIVFLDDRAQGSVNG	214
	orf3ng	IKEGISAQGEATMPPFAGNRKLAVIGAGGHGKVVAELAAALGTYGEIVFLDDRTQGSVNG	240
20	orf3	FSVIGTTLLLENSLSPEQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT	274
	orf3ng	FPVIGTTLLLENSLSPEQFDITVAVGNNRIRRQITENAAALGFKLPVLIHPDATVSPSAI	300
25	orf3	VGQGSVVMKAV	286
	orf3ng	IGQGSVVMKAVVQAGSVLKDGIVVNTAATVDHCLLDFAVHISPGAHLGNTTRIGEEER	360

The complete length ORF3ng nucleotide sequence <SEQ ID 17> is:

	1	ATGAGTAAAG	CCGTCAAACG	CCTGTTCGAC	ATCATCGCAT	CCGCATCGGG
	51	GCTGATTGTC	CTGTGCGCCG	TGTTTTTGGT	TTTAATATAC	CTCATCCGCA
30	101	AAAACCTAGG	TTGCGCCGTC	TTCTTCattC	GGGAACGCCc	cgGAAAGGAc
	151	ggaaaaacCTT	TTAAATGGT	CAAAATCCGT	TCCAtgcgcg	acgcgcttGA
	201	TTCAGACGGC	ATTCGCTGTC	CCGATAGCGA	ACGCCTGACC	GATTTCCGCA
	251	AAAAATTACG	CGCCACCACT	TTGGACGAAC	TTCTGAATT	ATGGAATGTC
	301	CTCAAAGGCG	AGATGAGCCT	GGTCGGCCCC	CGCCCGCTTT	TGATGCAGTA
35	351	TCTGCCGCTT	TACAACAAAT	TTCAAAACCG	CGGCCACGAA	ATGAAACCGG
	401	GCATTACCGG	CTGGGCGCAG	GTCAACGGGC	GCAACGCGCT	TTCGTGGGAC
	451	GAAAAGTTCT	CCTGCGATGT	TTGGTACACC	GACAATTTCA	GCTTTTGGCT
	501	GGATATGAAA	ATCCTGTTTC	TGACAGTCAA	AAAAGTCTTG	ATTAAAGAAG
	551	GCATTTCCGC	GCAAGGGGAA	GCCACCATGC	CCCCTTTCGC	GGGGAATCGC
40	601	AAACTCGCCG	TTATCGGCGC	GGGCGGACAC	GGCAAAGTCG	TTGCCGAGCT
	651	TGCCGCCGCA	CTCGGCACAT	ACGGCGAAAT	CGTTTTTCTG	GACGACCGCA
	701	CCCAAGGCAG	CGTCAACGGC	TTCCCCGTCA	TCGGCACGAC	GCTGCTGCTT
	751	GAAAACAGTT	TATCGCCCGA	ACAATTCGAC	ATCACCGTCG	CCGTCGGCAA
	801	CAACCGCATC	CGCCGCCAAA	TCACCGAAAA	CGCCGCCGCG	CTCGGCTTCA
45	851	AACTGCCCGT	TCTGATTTCAT	CCCGACGCGA	CCGTCTCGCC	TTCTGCAATA
	901	ATCGGACAAG	GCAGCGTCGT	AATGGCGAAA	GCCGTCGTAC	AGGCCGGCAG
	951	CGTATTGAAA	GACGGCGTGA	TTGTGAACAC	TGCCGCCACC	GTCGATCACG
	1001	ACTGCCTGCT	TGACGCTTTC	GtccaCATCA	GCCCCGGGCG	GCACCTGTCTG
	1051	GGCAACACGC	GTATCGGCGA	AGAAAGCCGG	ATAGGCACGG	GCGCGTGCAG
50	1101	CCGCCAGCAG	ACAACCGTCG	GCAGCGGGGT	TACCgcccGT	GCAGGGgcGG
	1151	TTATCGTATG	CGACATCCCG	GACGGCATGA	CCGTCGCGGG	CAACCCGGCA
	1201	AAGCCCTTA	CGGGCAAAAA	CCCCAAGACC	GGGACGGCAT	AA

This encodes a protein having amino acid sequence <SEQ ID 18>:

55	1	MSKAVKRLFD	IIASASGLIV	LSPVFLVLIY	LIRKNLGSPV	FFIRERPGKD
	51	GKPFKMKFR	SMRDALDSG	IPLPDSERLT	DFGKKLRATS	LDELPELWNV
	101	LKGEMSLVGP	RPLLMQYLPL	YNKFQNRHE	MKPGITGWAQ	VNGRNALSWD
	151	EKFSCDVWYT	DNFSFWLDMK	ILFLTVKKVL	IKEGISAQGE	ATMPPFAGNR
	201	KLAVIGAGGH	GKVVAELAAA	LGTYGEIVFL	DDRTQGSVNG	FPVIGTTLL
	251	ENSLSPEQFD	ITVAVGNNRI	RRQITENAAA	LGFKLPVLIH	PDATVSPSAI
60	301	IGQGSVVMK	AVVQAGSVLK	DGVIVNTAAT	VDHCLLDFA	VHISPGAHL
	351	GNTTRIGEE	IGTGACSRQ	TTVSGSVTAG	AGAVIVCDIP	DGMTVAGNPA
	401	KPLTGKNPK	GTA*			

This protein shows 86.9% identity in 413 aa overlap with ORF3-1:

		10	20	30	40	50	60
	orf3-1.pep	MSKFFKRLFDIVASASGLIFLSPVFLILYLRKKNL	GSPVFFFQ	ERPGKDGKPF	KMKVKFR		
5	orf3ng	MSKAVKRLFDIIASASGLIVLSPVFLVLIYLRKKNL	GSPVFFIR	ERPGKDGKPF	KMKVKFR		
		10	20	30	40	50	60
	orf3-1.pep	SMRDALDS	DGIPDPGERLT	PFGKKLRAASL	DELPELWNIL	KGEMSLV	GPRLMLQYLPL
10	orf3ng	SMRDALDS	DGIPDPGERLT	DGFKKLRAATSL	DELPELWNVL	KGEMSLV	GPRLMLQYLPL
		70	80	90	100	110	120
	orf3-1.pep	YDNFQNR	RHEMKPGIT	GWAQVNGRN	ALSWDEK	FACDVWY	IDHFSCLCDIKILL
15	orf3ng	YDNFQNR	RHEMKPGIT	GWAQVNGRN	ALSWDEK	FACDVWY	IDHFSCLCDIKILL
		130	140	150	160	170	180
	orf3-1.pep	IKEGISAQ	GEATMPFFT	GKRLAVV	GAGGHGKV	VADLAAAL	GRYREIVFLDDRAQ
20	orf3ng	IKEGISAQ	GEATMPFFT	GKRLAVV	GAGGHGKV	VADLAAAL	GRYREIVFLDDRAQ
		190	200	210	220	230	240
	orf3-1.pep	FSVIGT	TLLLENSL	SPEQYDV	AVAVGN	NNRIR	RQIAEKAAALGFALP
25	orf3ng	FSVIGT	TLLLENSL	SPEQYDV	AVAVGN	NNRIR	RQIAEKAAALGFALP
		250	260	270	280	290	300
	orf3-1.pep	VGQGSV	MAKAVVQ	AGSVL	KDGVIV	NTAATV	DHDCLLNAFVHIS
30	orf3ng	VGQGSV	MAKAVVQ	AGSVL	KDGVIV	NTAATV	DHDCLLNAFVHIS
		310	320	330	340	350	360
	orf3-1.pep	IGQGSV	MAKAVVQ	AGSVL	KDGVIV	NTAATV	DHDCLLDAFVHIS
35	orf3ng	IGQGSV	MAKAVVQ	AGSVL	KDGVIV	NTAATV	DHDCLLDAFVHIS
		370	380	390	400	410	
	orf3-1.pep	IGTGAC	SRQQIR	IGSRAT	IGAGAV	VVRD	VS
40	orf3ng	IGTGAC	SRQQIR	IGSRAT	IGAGAV	VVRD	VS

In addition, ORF3ng shows significant homology with a hypothetical protein from *B. subtilis*:

45	gnl PID e238668 (271928) hypothetical protein [Bacillus subtilis]
	>gi 1945702 gnl PID e313004 (294043) hypothetical protein [Bacillus subtilis]
	>gi 2635938 gnl PID e1186113 (299121) similar to capsular polysaccharide biosynthesis [Bacillus subtilis] Length = 202
	Score = 235 bits (594), Expect = 3e-61
	Identities = 114/195 (58%), Positives = 142/195 (72%)
50	Query: 5 VKRLFDIIASASGLIVLSPVFLVLIYLRKKNL
	+KRLFD+ A+ L S + L I ++R +GSPVFF + RPG GKPF + KFR+M D
	Sbjct: 3 LKRLFDLTAAIFLLCCTSVIILFTIAVVR
55	Query: 65 ALDS
	DS G LPD RLT G+ +R S+DELP+L NVLKG++SLVGPRLMLQYLPLY +
	Sbjct: 63 ERDSKGNLLPDEVRLTKTGR
60	Query: 125 QNRRHEMKPGITGWAQVNGRNALSWDEK
	Q RRHE+KPGITGWAQ+NGRNA+SW++KF DVWY DN+SF+LD+KIL LTV+KVL+ EG
	Sbjct: 123 QARRHEVKPGITGWAQINGRNAISWEKKFELDVWYVDNWSFFLDL
65	Query: 185 ISAQGEATMPFFAGN 199
	I T F G+
	Sbjct: 183 IQQTNHVTAERFTGS 197

The hypothetical product of *yvfc* gene shows similarity to EXOY of *R.meliloti*, an exopolysaccharide production protein. Based on this and on the two predicted transmembrane regions in the homologous *N.gonorrhoeae* sequence, it is predicted that these proteins, or their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 5 Example 4

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 19>:

```

1  ..AACCATATGG CGATTGTCAT CGACGAATAC GCGGCACAT CCGGCTTGTT
51  CACCTTTGAA GACATCATCG AGCAAATCGT CCGCGAAATC GAAGACGAGT
101 TTGACGAAGA CGATAGCGCC GACAATATCC ATGCCGTTTC TTCAGACACG
151 TGGCGCATCC ATGCAGCTAC CGAAATCGAA GACATCAACA CCTTCTTCGG
201 CACGGAATAC AGCATCGAAG AAGCCGACAC CATT.GGCGG CCTGGTCATT
251 CAAGAGTTGG GACATCTGCC CGTGCGCGGC GAAAAAGTCC TTATCGGCGG
301 TTTGCAGTTC ACCGTCGCAC GCGCCGACAA CCGCCGCCTG CATACGCTGA
351 TGGCGACCCG CGTGAAGTAA GC..... ACCGC CGTTTCTGCA
15 401 CAGTTTAG

```

This corresponds to amino acid sequence <SEQ ID 20; ORF5>:

```

1  ..NHMAIVIDEY GGTSGLVTFF DIIEQIVGEI EDEFDEDDSA DNIHAVSSDT
51  WRIHAATEIE DINTFFGTEY SIEEADTI XR PGHSRVGTSA RARRKSPYRR
101 FAVHRRTRRQ PPAAYADGDP REV.....XR RCTV*

```

20 Further sequence analysis revealed the complete DNA sequence to be <SEQ ID 21>:

```

1  ATGGACGGCG CACAACCGAA AACGAATTTT TTTGAACGCC TGATTGCCCCG
51  ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTAAAC CTGCTTCGGC
101 AGGCGCACGA GCAGGAAGTT TTTGATGCGG ATACGCTTTT AAGATTGGAA
151 AAAGTCCTCG ATTTTCCGA TTTGGAAGTG CGCGACGCGA TGATTACGGC
25 201 CAGCCGTATG AACGTTTAA AAGAAAACGA CAGCATCGAG CGCATCACCG
251 CCTACGTTAT CGATACCGCC CATTGCGGCT TCCCCGTCAT CCGCGAAGAC
301 AAAGACGAAG TTTTGGGCAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTTAACCCC GAGCAGTTCC ACCTCAAATC CATTCTCCGC CCCGCCGTCT
401 TCGTCCCCGA AGGCAAATCG CTGACCGCCC TTTTAAAGA GTTCCGCGAA
30 451 CAGCGCAACC ATATGGCGAT TGTCATCGAC GAATACGGCG GCACATCCGG
501 CTTGGTCACC TTTGAAGACA TCATCGAGCA AATCGTCGGC GAAATCGAAG
551 ACGAGTTTGA CGAAGACGAT AGCGCCGACA ATATCCATGC CGTTTCTTCC
601 GAACGCTGGC GCATCCATGC AGCTACCGAA ATCGAAGACA TCAACACCTT
651 CTTCCGACAG GAATACAGCA GCGAAGAAC CGACACCATT CGGCCTGGTC
35 701 ATTCAGAGT TGGGACATCT GCCCGTGCGC GGCGAAAAAG TCCTTATCGG
751 CGGTTTGCAG TTCACCGTCG CACGCGCCGA CAACCGCCGC CTGCATACGC
801 TGATGGCGAC CCGCGTGAAG TAAGCACCGC CGTTTCTGCA CAGTTTAGGA
851 TGACGGTACG GCGGTTTCT GTTCAATCC GCCCCATCCG CCAAACATAA

```

This corresponds to amino acid sequence <SEQ ID 22; ORF5-1>:

```

40 1  MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLRLLE
51  KVLDFSDLEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVLGILHA KDLLKYMFP EQFHLKSILR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSLVT FEDIIEQIVG EIEDEFDEDD SADNIHAVSS
201 ERWRIHAATE IEDINTFFGT EYSSEADTI RPHSRVGTSA ARARRKSPYR
45 251 RFAVHRRTRR QPPPAYADGD PREVSTAVSA QFRMTVRAFS VSIRPIRQT*

```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 23 >:

```

1  ATGGACGGCG CACAACCGAA AACAAATTTT TTNNAACGCC TGATTGCCCCG
51  ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTGACC CTGTTGCGCC
101 AAGCGCACGA ACAGGAAGTA TTTGATGCGG ATACGCTTTT AAGATTGGAA
50 151 AAAGTCCTCG ATTTTCTGTA TTTGGAAGTG CGCGACGCGA TGATTACGGC
201 CAGCCGTATG AACGTTTAA AAGAAAACGA CAGCATCGAA CGCATCACCG
251 CCTACGTTAT CGATACCGCC CATTGCGGCT TCCCCGTCAT CCGTGAAGAC
301 AAAGACGAAG TTTTGGGTAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTCAACCCC GAGCAGTTCC ACCTCAAATC GATATTGCGC CCTGCCGTCT

```

10

This encodes a protein having amino acid sequence <SEQ ID 24; ORF5a>:

15

20 The originally-identified partial strain B sequence (ORF5) shows 54.7% identity over a 124aa overlap with ORF5a:

25

```

                                10      20      30
orf5.pep                      NHMAIVIDEYGGTSGLVTFEDIIEQIVGEI
                                |||||
orf5a      FHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVGDI
              130      140      150      160      170      180

```

30

```

                                40      50      60      70      80      90
orf5.pep  EDEFDEDDSDADNIHAVSSDTWRIHAATEIEDINTFFGTEYSIEEADTIXRPGHSRVGTSA
          |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::||
orf5a     EDEFDEDESADNIHAVSAERWRIHAATEIEDINAFGGTEYSSEEADTIGGXGHSIGITPA
          190      200      210      220      230      240

```

35

```

              100      110      120      130
orf5.pep      RARRKSPYRRFAVHRRTRRQPPAYADGDPREVSVXXXXXRRFCTV
              ||||| ||| | | | : | ||||| |||||
orf5a         RARRKSXYRRXAXHXRXXQPPAYADGDPREVSSAVSVQFRMTVRAFSVSIRPIRXTX
              250      260      270      280      290      300

```

The complete strain B sequence (ORF5-1) and ORF5a show 92.7% identity in 300 aa overlap:

40

```

              10          20          30          40          50          60
orf5a.pep    MDGAQPKTNFXXRLIARLAREPDSAEDVLTLLRQAHEQEVFDADTLLRLEKVLDFSDEV
              |||||
orf5-1       MDGAQPKTNFFERLIARLAREPDSAEDVLNLLROAHEQEVFDADTLLRLEKVLDFSDEV

```

45

	10	20	30	40	50	60
orf5a.pep	70	80	90	100	110	120
	RDAMITRSRMNVLKENDS	IERITAYVIDTAHSRFP	VGIEDKDEV	GILHAKDLLK	YMFNP	
orf5-1	RDAMITRSRMNVLKENDS	IERITAYVIDTAHSRFP	VGIEDKDEV	GILHAKDLLK	YMFNP	
	70	80	90	100	110	120

50

0115-1	70	80	90	100	110	120
--------	----	----	----	-----	-----	-----

55

130            140            150            160            170            180  
 orf5a.pep    EQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIIEQIVG  
 |||||  
 orf5-1       EQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIIEQIVG  
 130            140            150            160            170            180

60

	190	200	210	220	230	240
orf5a.pep	DIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAF	FGTEYSSEEADTIGGXHGSHGIGT				
	:	:	:	:		:
orf5-1	EIEDEFDEDSADNIHAVSSERWRIHAATEIEDINT	FFGTEYSSEEADTIRP-GHSRVGT				
	190	200	210	220	230	
	250	260	270	280	290	300

orf5a.pep PARARRKSKYRRXAXHXRXRQPPPAYADGDPREVSSAVSQFRMTVRAFSVSIRPIRXT  
| | | | | | | | : | | | | | | | | : | | | | | | |  
orf5-1 SARARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSTAVSAQFRMTVRAFSVSIRPIRQT  
240 250 260 270 280 290

5 Further work identified the a partial DNA sequence in *N.gonorrhoeae* <SEQ ID 25> which encodes  
a protein having amino acid sequence <SEQ ID 26; ORF5ng>:

```

1 MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLTRLE
51 KVLDFAELEV RDAMITSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVLGILHA KDLLKYMENP EQFHLSKVLR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSGLVT FEDIEEQIVG DIEDEFDEDE SADDIHSVSA
201 ERWRIHAATE IEDINAFGT EYGSEADTI RRLGHSGIGT PARARRKSPY
251 RRFVAVHRRPR ROPPPAHADG DPREVSRACP HRHRECTV*

```

Further analysis revealed the complete gonococcal nucleotide sequence <SEQ ID 27> to be:

15	1	ATGGACGGCG	CACAACCGAA	AACAAATTTT	TTTGAACGCC	TGATTGCCCG
	51	ACTCGCCCGC	GAACCCGATT	CCGCCGAAGA	CGTATTAAC	CTGCTTCGGC
	101	AGGCGCACGA	ACAGGAA GTT	TTTGATGCCG	ACACACTGAC	CCGCTCGGAA
	151	AAAGTATTGG	ACTTTGCCGA	GCTGGAAGTG	CGCGATCGCA	TGATTACCGC
20	201	CAGCCGCATG	AACGTATTGA	AAGAAAACGA	CAGCATCGAA	CGCATACCCG
	251	CCTACGTCAT	CGATACCGCC	CATTGCGGCT	TCCCCGTCAT	CGGCGAAGAC
	301	AAAGACGAAG	TTTTGGGCAT	TTTGCACGCC	AAAGACCTGC	TCAAATATAT
	351	GTTCAACCCC	GAGCAGTTCC	ACCTGAAATC	CGTCTTGCGC	CCTGCCGTTT
25	401	TCGTGCCCGA	AGGCAAATCT	TTGACCGCCC	TTTAAAGA	GTCCGCGGAA
	451	CAGCGCAACC	ATATGGCAAT	CGTCATCGAC	GAATACGGCG	GCAGCTCGGG
	501	TTTGGTCACC	TTTGAAGACA	TCATCGAGCA	AATCGTCGGT	GACATCGAAG
	551	ACGAGTTTGA	CGAAGACGAA	AGCGcgcgacg	acatCCACTC	cgtTTccgCC
30	601	GAACGCTGGC	GCATCCacgc	ggctaCCGAA	ATCGAAGaca	TCAACGCCTT
	651	TTTCGGTACG	GAatacggca	gcgaagaagc	cgaccattc	cgccggttTG
	701	GTCATTACGG	AATTGGGACA	CCTGCCCGTG	CGCGGCGAAA	AAGTCCTTat
	751	cggcgGTTTG	Cagttcacgc	tCGCCCGCGC	CGACAACCGC	CGCTGCACA
30	801	CGTGATGCGC	GACCCGCGTG	AAGTAAGCAG	AGCTTGCCcg	AccgcggttT
	851	CTGCacAGTT	TAGGatgACG	gtaCGGTCTG	TTTCTGTTTC	AATCCGCCCC
	901	ATCCGCCAAA	CATAA			

This encodes a protein having amino acid sequence <SEQ ID 28; ORF5ng-1>:

```

35      1 MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLTRLE
      51 KVLDFAELEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
      101 KDEVLGILHA KDLLKYMFPN EQFHLKSVLR PAVFVPEGKS LTALLKEFRE
      151 QRNHMAIVID EYGGTSLVLT FEDIIEQIVG DIEDEFDEDE SADDIHSVSA
      201 ERWRIHAATE IEDINAFFGT EYGSEEDATI RRLGHSGIGT PARARRKSPY
40      251 RRFVAVHRRPR RQPPPAHADG DPREVSRACP TAVSAQFRMT VRSFSVSIRP
      301 IROT*

```

The originally-identified partial strain B sequence (ORF5) shows 83.1% identity over a 135aa overlap with the partial gonococcal sequence (ORF5ng):

	orf5	NHMAIVIDEYGGTSGLVTTFEDIIEQIVGEI	30
45	orf5ng	FHLKSVLRPAV FVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTTFEDIIEQIVGDI	182
	orf5	EDEFDEDDSDADNIHAVSSDTWRIHAATEIEDINTFFGTEYSIEEADTI XRPGHSRVGTS A	90
		:   :   :   :   :   :   :   :   :   :   :   :   :   :	
50	orf5ng	EDEFDEDESADDIHVSVAERWRIHAATEIEDINAFFGTEYGSEEDTIRRLGHSGIGTPA	242
	orf5	RARRKSPYRRFAVHRRTTRRQP PPAYADGDPREVSX----RRFCTV	131
	orf5ng	RARRKSPYRRFAVHRPRRRQP PAHADGDPREVRACPHRRFCTV	287

55 The complete strain B and gonococcal sequences (ORF5-1 & ORF5ng-1) show 92.4% identity in 304 aa overlap:

orf5nq-1.pep MDGAQPKTNFFERLIARLAREPDSAEDVLNLLROAEHQEVFDADTLTRLEKVLDFAELEV

**Homology with hemolysin homolog TlyC (accession U32716) of *H.influenzae***

```

ORF5      2      HMAIVIDEYGGTSGLVTFEDIIEQIVGEIEDEFDEDDSDADNIHAVSSDTWRIHAATEIED 61
           HMAIV+DE+G  SGLVT EDI+EQIVG+IEDEFDE++ AD I  +S  T+  + A T+I+D
TlyC     166     HMAIVVDEFGAVSGLVTIEDILEQIVGDIEDEFDEEEIAD-IRQLSRHTYAVRALTDIDD 224

ORF5      62     INTFFGTEYSIEEADTI 78
           N  F T++  EE DTI
TlyC     225     FNAQFNTDFDDEEVDTI 241

```

SCORES Init1: 301 Initn: 419 Opt: 668  
Smith-Waterman score: 668; 45.9% identity in 242 aa overlap

[illegible]



-73-

```

                    130      140      150      160      170      180
5      orf5ng-1.pep 170      180      190      200      210      220
                    VTFEDIIEQIVGDIIEFDEDESADDIHSVSAERWRIHAATEIEDINAFFGTEYGSSEAD
                    ||:||||:|||||||:| || |::| : : || |:|:|:| | |:|:|:|:|
tlyc_haein 170      180      190      200      210      220      230
                    VTIEDILEQIVGDIIEFDEEEIAD-IRQLSRHTYAVRALTDIDDFNAQENTDFDDEEVD
                    190      200      210      220      230

10     orf5ng-1.pep 230      240      250      260      270      280
                    TIRRLGHSGIG-TPARARRKSPYRRFAVHRRPRRQPPPAHADGDPREVSRACTAVSAQF
                    || | : :| | | :
tlyc_haein 230      240      250      260      270      280      290
                    TIGGLIMQTFGYLPKRGEIILKNLQFKVTSADSRRLIQLRVTVPDEHLAEMNNVDEKSE
                    240      250      260      270      280      290

```

### 15 Homology with a hypothetical secreted protein from *E.coli*:

ORF5a shows homology to a hypothetical secreted protein from *E.coli*:

```

20     sp|P77392|YBEX_ECOLI HYPOTHETICAL 33.3 KD PROTEIN IN CUTE-ASNB INTERGENIC REGION
>gi|1778577 (U82598) similar to H. influenzae [Escherichia coli] >gi|1786879
(AE000170) f292; This 292 aa ORF is 23% identical (9 gaps) to 272 residues of an
approx. 440 aa protein YTFL_HAEIN SW: P44717 [Escherichia coli] Length = 292

```

Score = 212 bits (533), Expect = 3e-54  
Identities = 112/230 (48%), Positives = 149/230 (64%), Gaps = 3/230 (1%)

```

25     Query: 2   DGAQPKTNFXRLIARLAR-EPDSAEDVLTLLRQAHEQEVFDADTLLRLEKVLDFSDLEV 60
                D   K   F   L+++L   EP + +++L L+R + + ++ D DT   LE V+D +D V
Sbjct: 10   DTISNKKGFFSLLLSQLFHGEPKRNDELLALIRDSGQNDLIEDTRDMLEGVMDIADQRV 69

30     Query: 61  RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYM-FN 119
                RD MI RS+M   LK N +++   +I++AHSRFPVI EDKD + GIL AKDLL +M +
Sbjct: 70   RDIMIPRSQMITLKRNTLDECLDVIIESAHSRFPVISEDKDHIEGILMAKDLLPFMRSD 129

35     Query: 120 PEQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIV 179
                E F + +LR AV VPE K + +LKEFR QR HMAIVIDE+GG SGLVT EDI+E IV
Sbjct: 130  AEAFSMDKVLQRAVVVPESKRVRMLKEFRSQRYHMAIVIDEFGGVSGLVTFIEDILELIV 189

        Query: 180 GDIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFFGTEYSSEEADT 229
                G+IEDE+DE++ D   +S   W + A   IED N   FGT +S EE DT
Sbjct: 190  GEIEDEYDEEDDID-FRQLSRHTWTVRALASIEDFNEAFGTHFSDEEVD 238

```

40 Based on this analysis, including the amino acid homology to the TlyC hemolysin-homologue from *H. influenzae* (hemolysins are secreted proteins), it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae* are secreted and could thus be useful antigens for vaccines or diagnostics.

45 ORF5-1 (30.7kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 2A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot analysis (Figure 1B). These experiments confirm that ORF5-1 is a surface-exposed protein, and that it is a useful immunogen.

### Example 5

50 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 29>:

```

1   ATGCGCGGCG GCAGGCCGGA TTCCGTTACC GTGCAGATTA TCGAAGGTTG
51  GCGTTTTCG CATATGAGGA AAGTCATCGA CGCAACGCC GACATCGGAC

```

5  
10  
101 ACGACACCAA AGGCTGGAGC AATGAAAAAC TGATGGCGGA AGTTGCGCCC  
151 GATGCCTTCA GCGGCAATCC TGAAGGCGAG TTTTCCCGG ACAGTACGA  
201 AATCGATGCG GCGGCGAGT ATTTGCAGAT TTACCAAACC GCCTACAAgG  
251 GCGATGCAAC GCCGCCTGAA TGAAGGCGATG GGAAGCAGG CAGGACGGGC  
301 TGCCTTATAA AAACCCTTAT GAAATGCTGA TTATGGCGAr CCTGGTCGAA  
351 AAGGAAACAG GGCATGAAGC CGAsCsCGAC CATGTcGCTT CCGTCTTCGT  
401 CAACCGCCTG AAAATCGGTA TCGCCTGCA AACCgAssCG TCCGTGATTT  
451 ACGGCATGGG TCGGCATAC AAGGGCAAAA TCCGTAAAGC CGACCTGCGC  
501 CGCGACACGC CGTACAACAC CTACACGCGC GCGGGTCTGC CGCCAACCCC  
551 GATTGCGCTG CCC..

This corresponds to the amino acid sequence <SEQ ID 30; ORF7>:

15  
1 MRGGRPDSVT VQIIEGSRFS HMRKVIDATP DIGHDTKGWS NEKLMAEVAP  
51 DAFSGNPEGQ FFPDSYEIDA GGS DLQIYQT AYKAMQRRNL EAWESRDGL  
101 PYKNPYEMLI MAXLVEKETG HEAXDXHVAS VFNRLKIGM RLQTXXSIVI  
151 GMGAAYKGI RKADLRDTP YNTYTRGGLP PTPIALP..

Further sequence analysis revealed the complete DNA sequence <SEQ ID 31>:

20  
25  
30  
35  
1 ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTTGA CCGTGTCCGG  
51 AGCCGTTTTT CCGCGCTGCT TTTTGTTC TAAGGATAAC GGCAGGGCAT  
101 ACCGAATCAA AATTGCCAAA AACCAGGGTA TTTCTCGGT CCGCAGGAAA  
151 CTTGCCGAAG ACCGCATCGT GTTACGAGG CATGTTTTGA CCGCGCGGC  
201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGGACG TACAGATTGC  
251 CTTCCGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CCGCGGCAGG  
301 CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTCCGCTT TTTTCGCATAT  
351 GAGGAAAGTC ATCGACGCAA CGCCCGACAT CGGACACGAC ACCAAAGGCT  
401 GGAGCAATGA AAAACTGATG GCGGAAGTTG CGCCCGATGC CTTACGCGGC  
451 AATCCTGAAG GGCAGTTTTT CCCCACAGC TACGAAATCG ATGCGGGCGG  
501 CAGTGATTG CAGATTTACC AAACCGCCTA CAAGGCGATG CAACGCGGCC  
551 TGAATGAGGC ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCT  
601 TATGAAATGC TGATTATGGC GAGCCTGGTC GAAAAGGAAA CAGGGCATGA  
651 AGCCGACCGC GACCATGTCG CTTCCGTCTT CGTCAACCGC CTGAAAATCG  
701 GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGGTGCGGCA  
751 TACAAGGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTACAA  
801 CACCTACACG CGCGCGGTC TGCCGCCAAC CCCGATTGCG CTGCCCGGCA  
851 AGCGGCACT CGATCCGCC GCCCATCCGT CCGGCGAAAA ATACCTGTAT  
901 TTCGTGTCCA AAATGGACGG CACGGGCTTG AGCCAGTTCA GCCATGATTT  
951 GACCGAACAC AATGCCGCG TCCGCAATA TATTTTGAAA AAATAA

This corresponds to the amino acid sequence <SEQ ID 32; ORF7-1>:

40  
1 MLRKLLKWSA VELTVSAAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK  
51 LAEDRIVFSR HVLTAAYVL GVHNRLHTGT YRLPSEVS AW DILQMRGGR  
101 PDSVTVQIIE GSRFSHMRKV IDATPDIGHD TKGWSNEKLM AEVAPDAFSG  
151 NPEGQFFPDS YEIDAGGSDL QIYQTAYKAM QRRLEAWES RQDGLPYKNP  
201 YEMLIMASLV EKETGHEADR DHVASVFVNR LKIGMRLQTD PSVIYGMGA  
251 YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGKAALDAA AHPSGEKYL  
301 FVSKMDGTGL SQFSDLTEH NAAVRKYILK K\*

45 Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein encoded by yceg gene (accession P44270) of *H. influenzae*

ORF7 and yceg proteins show 44% aa identity in 192 aa overlap:

50  
55  
ORF7 1 MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMA-----EVAPDAFSG 55  
+ G+ V+ IEG F RK ++ P + K SNE++ A ++ +  
yceg 102 LNSGKEVQFNVKWIEGKTFKDWKRDLENAPHLVQTLKDKSNEEIFALLDLPDIGQNLLEK 161  
ORF7 56 NPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWESRDGLPYKNPYEMLIMAXLV 115  
N EG +PD+Y +DL++ + + M++ LN+AW R + LP NPYEMLI+A +V  
yceg 162 NVEGWLYPDYNTYTPKSTDLLELLKRSERMKKALNKAWNERDEDLPLANPYEMLILASIV 221  
ORF7 116 EKETGHEAXDXHVASVFVFNRLKIGMRLQTXXSIVIYGMGAAYKGIKIRKADLRDTPYNTYT 175  
EKETG VASVF+NRLK M+LQT +VIYGMG Y G IRK DL TPYNTY  
yceg 222 EKETGIANERAKVASVFVFNRLKAKMKLQTDPTVIYGMGENYNGNIRKKDLETCTPYNTYV 281

ORF7 176 RGGLPPTPIALP 187  
GLPPTPIA+P  
yceg 282 IDGLPPTPIAMP 293

The complete length YCEG protein has sequence:

```

5      1 MKKFLIAILL LILILAGVAS FSYYKMTEFV KTPVNVQADE LLTIERGTTS
      51 SKLATLFEQE KLIADGKLLP YLLKLKPELN KIKAGTYSLE NVKTVQDLLD
     101 LLNSGKEVQF NVKWIEGKTF KDWRKDLENA PHLVQTLKDK SNEEIFALLD
     151 LPDIGQNLEL KNVEGWLYPD TYNYPKSTD LELLKRSAER MKKALNKAWN
     201 ERDEDPLAN PYEMLILASI VEKETGIANE RAKVASVFIN RLKAKMKLQT
10    251 DPTVIYGMGE NYNGNIRKRD LETKTPYNTY VIDGLPPTPI AMPSESSLQA
     301 VANPEKTDYF YFVADGSGGH KFTRNLNEHN KAVQEYLRWY RSQKNKAK

```

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF7 shows 95.2% identity over a 187aa overlap with an ORF (ORF7a) from strain A of *N.*

15 *meningitidis*:

```

                                     10      20      30
or7.pep                               MRGGRPDSTVTVQIIIEGSRFSHMRKVIDATP
                                     |||||
20    orf7a    AAYVLGVHNRLLHTGTYRLPSEVSAWDILQKMRGGRPDSTVTVQIIIEGSRFSHMRKVIDATP
               70      80      90      100     110     120

               40      50      60      70      80      90
or7.pep    DIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLN
25    orf7a    DIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAMQRRLN
               130     140     150     160     170     180

               100     110     120     130     140     150
or7.pep    EAWESRQDGLPYKNPYEMLIMAXLVEKETGHEAXXDHVASVFVNRLKIGMRLQTXSVIY
30    orf7a    EAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFVNRLKIGMRLQTDPSVIY
               190     200     210     220     230     240

               160     170     180
or7.pep    GMGAAYKKGKIRKADLRRDTPYNTYTRGGLPPTPIALP
35    orf7a    GMGAAYKKGKIRKADLRRDTPYNTYTRGGLPPTPIALPGKAALDAAHPSGEKYLYFVSKM
               250     260     270     280     290     300

40    orf7a    DGTGLSQFSHDLTEHNAAVRKYLKXX
               310     320     330

```

The complete length ORF7a nucleotide sequence <SEQ ID 33> is:

```

      1 ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTTGA CCGTATCGGC
     51 AGCCGTTTTC GCCGCGCTGC TTTTCGTCCC TAAAGACAAC GGCAGGGCAT
45    101 ACAGGATTAA AATTGCCAAA AACCAGGGTA TTTTCGTGGT CCGCAGGAAA
     151 CTTGCCGAAG ACCGCATCGT GTTCAGCAGG CATGTTTTGA CCGCGGCGGC
     201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGGACG TACAGACTGC
     251 CTTCCGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGGCGGCAGG
     301 CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTCCGCTT TTTTCGCATAT
50    351 GAGGAAAGTC ATCGACGCAA CGCCCGACAT CGAACACGAC ACCAAAGGCT
     401 GGAGCAATGA AAAACTGATG GCGGAAGTTG CCCCTGATGC CTTTCAGCGGC
     451 AATCCTGAAG GGCAGTTTTT CCCCACGACG TACGAAATCG ATGCGGGCGG
     501 CAGCGATTTA CGGATTACC AAATCGCCTA CAAGGCGATG CAACGCCGAC
     551 TGAATGAGGC ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCCT
55    601 TATGAAATGC TGATTATGGC GAGCCTGATC GAAAAGGAAA CAGGGCATGA
     651 AGCCGACCGC GACCATGTGG CTTCCGTCTT CGTCAACCGC CTGAAAATCG
     701 GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGGTGCGGCA
     751 TACAAGGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTACAA
     801 CACCTACACG CGCGGCGGTC TGCCGCCAAC CCGGATCGCG CTGCCCGGCA
60    851 AGGCGGCACT CGATGCCGCC GCCCATCCGT CCGGTGAAAA ATACCTGTAT
     901 TTCGTGTCCA AAATGGACGG TACGGGCTTG AGCCAGTTCA GCCATGATTT
     951 GACCGAACAC AACGCCGCGG TTCGCAATA TATTTTGAAA AAATAA

```

This is predicted to encode a protein having amino acid sequence <SEQ ID 34>:

```

      1 MLRKLLKWSA VFLTVSAAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK
    51 LAEDRIVFSR HVLTAAYVL GVHNRLHTGT YRLPSEVSAW DILQKMRGGR
   101 PDSVTVQIIE GSRFSHMRKV IDATPDIEHD TKGWSNEKLM AEVAPDAFSG
   151 NPEGQFFPDS YEIDAGGSDL RIYQIAYKAM QRRLEAWES RQDGLPYKNP
   201 YEMLIMASLI EKETGHEADR DHVASVFVNR LKIGMRLQTD PSVIYGMGAA
   251 YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGKAALDAA AHPSGEKYL
   301 FVSKMDGTGL SQFSHDLTEH NAAVRKYILK K*

```

A leader peptide is underlined.

10 ORF7a and ORF7-1 show 98.8% identity in 331 aa overlap:

```

      10      20      30      40      50      60
  orf7a.pep MLRKLLKWSAVFLTVSAAVFAALLFVPKDNGRAYRIKIAKNQGISSVGRKLAEDRIVFSR
            |||
  orf7-1    MLRKLLKWSAVFLTVSAAVFAALLFVPKDNGRAYRIKIAKNQGISSVGRKLAEDRIVFSR
            |||
      70      80      90     100     110     120
  orf7a.pep HVLTAAYVLGVHNRLHTGT YRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKV
            |||
  orf7-1    HVLTAAYVLGVHNRLHTGT YRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKV
            |||
      70      80      90     100     110     120
  orf7a.pep IDATPDIEHDTKGWSNEKLM AEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAM
            |||
  orf7-1    IDATPDIGHDTKGWSNEKLM AEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAM
            |||
      130     140     150     160     170     180
  orf7a.pep QRRLEAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFVNR LKIGMRLQTD
            |||
  orf7-1    QRRLEAWESRQDGLPYKNPYEMLIMASLVEKETGHEADRDHVASVFVNR LKIGMRLQTD
            |||
      190     200     210     220     230     240
  orf7a.pep PSVIYGMGAAYKKGKIRKADLRRDTPYNTYTRGGLPPTPIALPGKAALDAAAHPSGEKYL
            |||
  orf7-1    PSVIYGMGAAYKKGKIRKADLRRDTPYNTYTRGGLPPTPIALPGKAALDAAAHPSGEKYL
            |||
      250     260     270     280     290     300
  orf7a.pep FVSKMDGTGLSQFSHDLTEHNAAVRKYILKXX
            |||
  orf7-1    FVSKMDGTGLSQFSHDLTEHNAAVRKYILKXX
            |||
      310     320     330
  orf7a.pep
            |||
  orf7-1
            |||

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF7 shows 94.7% identity over a 187aa overlap with a predicted ORF (ORF7.ng) from *N. gonorrhoeae*:

```

   50  orf7      MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLM AEVAPDAFSGNPEGQ 60
      orf7ng   MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLM AEVAPDAFSGNPEGQ 60
   55  orf7      FFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWESRQDGLPYKNPYEMLIMAXLVEKETG 120
      orf7ng   FFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWAGRQDGLPYKNPYEMLIMASLIEKETG 120
      orf7      HEAXXDHVASVFVNR LKIGMRLQTXXSVIYGMGAAYKKGKIRKADLRRDTPYNTYTRGGLP 180
      orf7ng   HEADRDHVASVFVNR LKIGMRLQTDPSVIYGMGAAYKKGKIRKADLRRDTPYNTYTGGGLP 180
      orf7      PTPIALP

```

or7ng                    || ||||  
 PTRIALPGKAAMDAAAHPSGEKYLYFVSKMDGTGLSQFSHDLTEHNAAVRKYILKK 236

An ORF7ng nucleotide sequence <SEQ ID 35> is predicted to encode a protein having amino acid sequence <SEQ ID 36>:

5                    1 MRGGRPDSVT VQIEGSRFS HMRKVIDATP DIGHDTKGWS NEKLMAEVAP  
                   51 DAFSGNPEGQ FFPDSYEIDA GGSDLQIYQT AYKAMQRRLN EAWAGRQDGL  
                   101 PYKNPYEMLI MASLIEKETG HEADRDHVAS VFNRLKIGM RLQTDPSVIY  
                   151 GMGAAYKGKI RKADLRDTP YNTYTGGGLP PTRIALPGKA AMDAAAHPSG  
                   201 EKLYFVSKM DGTGLSQFSH DLTEHNAAVR KYILKK\*

10 Further sequence analysis revealed a partial DNA sequence of ORF7ng <SEQ ID 37>:

                  1 ..taccgaatca AGATTGCCAA AAATCAGGGT ATTCGTCGG TCGGCAGGAA  
                   51 ACTTGCcgaA GACCGCATCG TGTTTCAGCAG GCATGTTTGT ACAGCGGCGG  
                   101 CCTACGTTTT GGGTGTGCAC AACAGGCTGC ATACGGGGAC gTACAGATG  
                   151 CCTTCGGAAG TGCTGCTTG GGATATCTTG CAGAAAATGC GCGGCGGCAG  
 15                    201 GCCGGATTCC GTTACCGTGC AGATTATCGA AGGTTTCGGT TTTTCGCATA  
                   251 TGAGGAAAGT CATCGACGCA ACGCCCGACA TCGGACACGA CACCAAAGGC  
                   301 TGGAGCAATG AAAAATGAT GCGGGAAGTT GCGCCGATG CCTTCAGCGG  
                   351 CAATCCTGAA GGCAGTTTTT TTCCCGACAG CTACGAAATC GATGCGGCGG  
                   401 GCAGCGATTT GCAGATTTAC CAAACCGCCT ACAAGGCGAT GCAACGCCGC  
 20                    451 CTGAACGAGG CATGGGCAGG CAGGCAGGAC GGGCTGCCTT ATAAAAACCC  
                   501 TTATGAAATG CTGATTATGG CGAGCCTGAT CGAAAAGGAA ACGGGGCATG  
                   551 AGGCCGACCG CGACCATGTC GCTTCCGTCT TCGTCAACCG CCTGAAAATC  
                   601 GGTATGCGCC TGCAAACCGA CCCGTCCTG ATTTACGGCA TGGGTGCGGC  
                   651 ATACAAGGGC AAAATCCGTA AAGCCGACCT GCGCCGCGAC ACGCCGTACA  
 25                    701 aCAcTAtac gggcgggggc ttgccgcaa cccgattgc gctgcccggC  
                   751 Aagcgggcaa tggatgccgc cgcccaccg tccgcgcaAa aatacctgTa  
                   801 tttcgtgtcC AAAATGGACG GCACGGGCTT GAGCCAGTTC AGCCATGATT  
                   851 TGACCGAACA CAACGCCGcC gTcCGCAAT ATATTTTGAA AAAATAA

This corresponds to the amino acid sequence <SEQ ID 38; ORF7ng-1>:

30                    1 ..YRIKIAKNQG ISSVGRKLAE DRIVFSRHVL TAAAYVLGVH NRLHTGTYRL  
                   51 PSEVSAWDIL QKMRGGRPDS VTVQIEGSR FSHMRKVIDA TPDIGHDTKG  
                   101 WSNKLMAEV APDAFSGNPE GQFFPDSYEI DAGGSDLQIY QTAYKAMQRR  
                   151 LNEAWAGRQD GLPYKNPYEM LIMASLIEKE TGHEADRDHV ASVFNRLKI  
 35                    201 GMRLQTDPSV IYGMGAAYKG KIRKADLRD TPYNTYTGGG LPPTRIALPG  
                   251 KAAMDAAHP SGEKYLYFVS KMDGTGLSQF SHDLTEHNAA VRKYILKK\*

ORF7ng-1 and ORF7-1 show 98.0% identity in 298 aa overlap:

	10	20	30	40	50	60
orf7-1.pep	KLLKWSAVFLTVSAAVFAALLFVPKDNGRAYRIKIAKNQGISSVGRKLAE DRIVFSRHVL					
orf7ng-1	YRIKIAKNQGISSVGRKLAE DRIVFSRHVL					
	10	20	30	40	50	60
orf7-1.pep	TAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIEGSRFSHMRKVIDA					
orf7ng-1	TAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIEGSRFSHMRKVIDA					
	40	50	60	70	80	90
orf7-1.pep	TPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRR					
orf7ng-1	TPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRR					
	100	110	120	130	140	150
orf7-1.pep	LNEAWESRQDGLPYKNPYEMLIMASLVEKETGHEADRDHVASVFNRLKIGMRLQTDPSV					
orf7ng-1	LNEAWAGRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFNRLKIGMRLQTDPSV					
	160	170	180	190	200	210
orf7-1.pep	IYGMGAAYKGKIRKADLRDTPYNTYTGGGLPPTPIALPGKAALDAAAHPSGEKYLYFVS					

```

|||||
orf7ng-1  IYGMGAAYKGKIRKADLRDTPYNTYTGGLPPTRIALPGKAAMDAAAHPSGEKILYFVS
          220      230      240      250      260      270
          310      320      330
orf7-1.pep KMDGTGLSQFSHDLTEHNAAVRKYILKKX
          |||||
orf7ng-1   KMDGTGLSQFSHDLTEHNAAVRKYILKKX
          280      290

```

In addition, ORF7ng-1 shows significant homology with a hypothetical *E.coli* protein:

sp|P28306|YCEG\_ECOLI HYPOTHETICAL 38.2 KD PROTEIN IN PABC-HOLB INTERGENIC REGION  
gi|1787339 (AE000210) o340; 100% identical to fragment YCEG\_ECOLI SW: P28306 but  
has 97 additional C-terminal residues [Escherichia coli] Length = 340  
Score = 79 (36.2 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57  
Identities = 20/87 (22%), Positives = 40/87 (45%)

Query: 10 GISSVGRKLAEDRIVFSRHLVLTAAAYVLGVHNRLLHTGTYRLPSEVSAWDILQKMRGGRPD 69  
G ++G +L D+I+ V + + GTYR +++ ++L+ + G+

Subject: 49 GRLALGEQLYADKIINRPRVFQWLLRIEPLSHFKAGTYRFTPMQTVREMLKLLESGKEA 108

```
Query:      70 SVTVQIIIEGSRFSHMRKVIDATPDIGH 96
           +++++EG R S   K +   P I H
Spict:     109 QFPLRLVEGMRLSDYLKQLREAPYIKH 135
```

Score = 438 (200.7 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57  
Identities = 84/155 (54%), Positives = 111/155 (71%)

Query: 120 EGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRRLNEAWAGRQDGLPYKNPYEMLIMASLIEK 179  
EG F+PD++ A +D+ + + A+K M + ++ AW GR DGLPYK+ +++ MAS+IEK  
Sbjct: 158 EGWFWPDTWMYTANTTDVALLKRAHKMKVKAUDSAWEGRADGLPYKDKNQLVTMASIIIEK 217

Query: 180 ETGHEADRDHVASVFVNRLKIGMRLQTDPSVIYGMGAAYKGKIRKADLRDTPYNTYTGG 239  
ET ++RD VASVF+NRL+IGMRLQTD+VIYGMG Y GK+ +ADL T YNTYT  
Sbjct: 218 ETAVASERDKVASVFINRLRIGMRLQTDPTVIYGMGERYNGKLSRADLETPTAYNTYTIT 277

Query: 240 GLPPTRIALPGKAAMDAAHPSGKEYLYFVSKMDG 274  
GLPP IA PG ++ AAAHP+ YLYFV+ G  
Sbjct: 278 GLPPGAIATPGADSLKAAAHPAKTPYLYFVADGKG 312

Based on this analysis, including the fact that the *H.influenzae* YCEG protein possesses a possible leader sequence, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 6

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 39>:

1	CGTTTCAAAA	TGTTAACTGT	GTTGACGGCA	ACCTTGATTG	C CGGACAGGT
51	ATCTGCCGCC	GGAGGCGGTG	CGGGGGATAT	GAAACAGCCG	AAGGAAAGTCG
101	GAAAGGTTTT	CAGAAAGCAG	CAGCGTTACA	GCGAGGAAGA	AATCAAAAAC
151	GAACCGCGAC	GGCTTGCGGC	AGTGGGCGAG	CGGGTTAATC	AGATATTTAC
201	GTTGCTGGGA	GGGGA AACCG	CGTTGCAAAA	GGGGCAGCGC	GGAACGGCTC
251	TGGCAACCTA	TATGCTGATG	TTGGAACGCA	CAAAATCCCC	CGAAGTCGCC
301	GAACGCGCCT	TGGAAATGGC	CGTGTGCTG	AACGCGTTTG	AACAGGCGGA
351	AATGATTTTG	CAGAAATGGC	GGCAGATTGA	GCCTATACCG	GGTAAGGCGC
401	AAAAACGGGC	GGGGTGGCTG	CGGAACGTGC	TGAGGGAACG	AGGAAATCAG
451	CAATCTGGACG	GACGGGAAGA	AGTGCTGGCT	CAGGCGGAGC	AAGGCACAG

This corresponds to the amino acid sequence <SEQ ID 40; ORF9>:

1 ..RFKMLTVLTA TLIAGQVSAA GGGAGDMKQP KEVGKVFVRKQ QRYSEEEIKN  
51 ERARLAAVGE RVNQIFTLTG GETALQKGQA GTALATYMLM LERTKSPEVA  
101 ERALEMAVSL NAFEQAEMIIY QKWROIPIPI GKAKRAGWL RNVLRERGNO

151 HLDGREEVLA QADEGO

Further sequence analysis revealed the complete DNA sequence <SEQ ID 41>:

	1	ATGTTACCTA	ACCGTTTCAA	AATGTTAACT	GTGTTGACGG	CAACCTTGAT
5	51	TGCCGGACAG	GTATCTGCCG	CCGGAGGCGG	TGCGGGGGAT	ATGAAACAGC
	101	CGAAGGAACT	CGGAAAGGTT	TTCAGAAAGC	AGCAGCGTTA	CAGCGAGGAA
	151	GAAATCAAAA	ACGAACGCGC	ACGGCTTGCG	GCAGTGGCGG	AGCGGGTTAA
	201	TCAGATATTT	ACGTTGCTGG	GAGGGGAAAC	GCGCTTGCAA	AAGGGGCGAG
10	251	CGGGAACGCG	TCTGGCAACC	TATATGCTGA	TGTTGGAACG	CACAAAATCC
	301	CCCGAAGTCG	CCGAACGCGC	CTTGGAATATG	GCCGTGTGCG	TGAACGCGTT
	351	TGAACAGGCG	GAAATGATTT	ATCAGAAATG	GCGGCAGATT	GAGCCTATAC
	401	CGGGTAAGGC	GCAAAAACGG	GCGGGGTGGC	TGCGGAACGT	GCTGAGGGAA
15	451	AGAGGAAATC	AGCATCTGGA	CGGACTGGAA	GAAGTGCTGG	CTCAGGCGGA
	501	CGAAGGACAG	AACCGCAGGG	TGTTTTTATT	GTTTGCACAA	GCCGCCGTGC
	551	AACAGGACGG	GTTGGCGCAA	AAAGCATCGA	AAGCGGTTCT	CCGCGCGGCG
	601	TTGAAATATG	AACATCTGCC	CGAAGCGGCG	GTTGCCGATG	TGTTGTTTCA
20	651	CGTACAGGGA	CGCGAAAAGG	AAAAGGCAAT	CGGAGCTTTG	CAGCGTTTGG
	701	CGAAGCTCGA	TACGGAAATA	TTGCCCCCA	CTTTAATGAC	TTGTCGCTCG
	751	ACTGCACGCA	AATATCCCGA	AATACTCGAC	GCTTTTTCG	AGCAGACAGA
	801	CACCCAAAC	CTTTCGGCCG	TCTGGCAGGA	AATGGAAATT	ATGAATCTGG
25	851	TTTCCCTGCA	CAGGCTGGAT	GATGCCTATG	CGCPTTGA	CGTCTGTTG
	901	GAACGCAATC	CGAATGCAGA	CCTGTATATT	CAGGCAGCGA	TATTGGCGGC
	951	AAACCGAAAA	GAAGGTGCTT	CCGTTATCGA	CGGCTACGCC	GAAGAGGCAT
	1001	ACGGCAGGGG	GACGGAGGAA	CAGCGGAGCA	GGCGCGGCT	AACGGCGGCG
30	1051	ATGATGTATG	CCGACCGCAG	GGATTACGCC	AAAGTCAGGC	AGTGGCTGAA
	1101	AAAAGTATCC	GCGCCGGAAT	ACCTGTTTCA	CAAAGGTGTG	CTGGCGGCTG
	1151	CGGCGGCTGT	CGAGTTGGAC	GGCGGCAGGG	CGGCTTTGCG	GCAGATCGGC
	1201	AGGGTGCGGA	AACTTCCCGA	ACAGCAGGGG	CGGTATTTTA	CGGCAGACAA
35	1251	TTTGTCCAAA	ATACAGATGC	TCGCCCTGTC	GAAGCTGCCC	GATAAACCGG
	1301	AGGCTTTAG	GGGGTTGGAG	AAGATTATCG	AAAAACCGCC	TGCCGCGAGT
	1351	AATACAGAGT	TACAGGCAGA	GGCATTGGTA	CAGCGGTCAG	TTGTTTACGA
	1401	TCGGCTTGCG	AAGCGGAAAA	AAATGATTTC	AGATCTTGAA	AGGGCGTTCA
40	1451	GGCTTGCACC	CGATAACGCT	CAGATTATGA	ATAATCTGGG	CTACAGCCTG
	1501	CTGACCGATT	CCAAACGTTT	GGACGAAGGT	TTCCGCTTGC	TTACAGACGG
	1551	ATACCAATC	AACCCGGACG	ATACCGCTGT	CAACGACAGC	ATAGGCTGGG
	1601	CGTATTACCT	GAAAGGCGAC	GCGGAAAGCG	CGCTGCCGTA	TCTGCGGTAT
45	1651	TCGTTTGA	ACGACCCCGA	GCCCCAAGTT	GCCGCCCATT	TGGGCGAAGT
	1701	GTTGTGGGCA	TTGGGCGAAC	GCGATCAGGC	GGTTGACGTA	TGGACGCGAG
	1751	CGGCACACCT	TACGGGAGAC	AAGAAAATAT	GGCGGGAAC	GCTCAAACGT
	1801	CACGGCTCGT	CATTGCCCCA	ACCTTCCCGA	AAACCTCGGA	AATAA

40 This corresponds to the amino acid sequence <SEQ ID 42; ORF9-1>:

	1	MLPNRFKMLT	VLATATLIAGO	VSAAGGGAGD	MKQPKVEGVK	FRKQORYSEE
	51	EIKNERARLA	AVGERVNQIF	TLLGGETALQ	KGQAGTALAT	YMLMLERTKS
	101	PEVAERALEM	AVSLNAFEQA	EMIYQKWRQI	EPIPGKAQKR	AGWLRNVLRE
45	151	RGNOHLDLGE	EVLQADEGQ	NRRVFLLLAQ	AAVQKDGLAQ	KASKAVRRAA
	201	LKYEHLPEAA	VADVVFVQGG	REKEKAIGAL	QRLAKLDTEI	LPPTLMTLRL
	251	TARKYPEILD	GFFEQTDTQN	LSAVWQEMEI	MNLVSLHRLD	DAYARLNVLL
	301	ERNPNADLYI	QAAAILAANRK	EGASVIDGYA	EKAYGRGTGE	QRSRAALTAQ
	351	MMYADRRDYA	KVRQWLKKS	APeyLFDKGV	LAAAARVELD	GGRAALRQIG
	401	RVRKLPEQQG	RYFTADNLSK	IQMLALSKLP	DKREALRGLD	KIIEKPPAGS
50	451	NTELQAEALV	QRSVVYDRIG	KRKKMISDLE	RAFRLAPDNA	QIMNNLGYSL
	501	LTDSKRLDEG	FALLQTAYQI	NPDDTAVNDS	IGWAYYILKGD	AESALPYLRY
	551	SFENDPEPEV	AAHLGEVLWA	LGERDQAVDV	WTQAAHLTGD	KKIWRET LKR
	601	HGIALPOPSR	KPRK*			

Computer analysis of this amino acid sequence gave the following results:

55 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF9 shows 89.8% identity over a 166aa overlap with an ORF (ORF9a) from strain A of *N. meningitidis*:

```

                                10      20      30      40      50
60 orf9.pep    RFKMLTVLTATLIAGQVSAAGGGAGDMKPKEVGKVFRKKQORYSEEEIKNERARLA
        || :||:|||:||||: ||   ||:| | ||| ||||| ||||| ||||| ||||| |||||
orf9a     MLPARETTLSVLAAALLAGQAYAA--GAADAKPPKEVGKVFRKKQORYSEEEIKNERARLA

```

-80-

		10	20	30	40	50	
		60	70	80	90	100	110
5	orf9.pep	AVGERVNQIFTLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA					
	orf9a	AVGERVNQIFTLGXETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA					
		60	70	80	90	100	110
		120	130	140	150	160	
10	orf9.pep	EMIQKWRQIEPIPGKAQKRAGWLRNVLRRGNQHLDGREEVLAQADEGQ					
	orf9a	EMIQKWRQIEPIPGKAQKRAGWLRNVLRRGNQHLDGLEEXLAQADEXQNRVFLLLAQ					
		120	130	140	150	160	170
15	orf9a	AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVFSVQXREKEKAIGALQRLAKLDTEI					
		180	190	200	210	220	230

The complete length ORF9a nucleotide sequence <SEQ ID 43> is:

	1	ATGTTACCG	CCCGTTTCAC	CATTTTATCT	GTGCTCGCGG	CAGCCCTGCT
	51	TGCCGGGCAG	GCGTATGCCG	CCGGCGCGGC	GGATGCGAAG	CCGCCGAAGG
20	101	AAGTCGGA	GGTTTTAGA	AAGCAGCAGC	GTTACAGCGA	GGAAGAAATC
	151	AAAAACGAAC	GCGCACGGCT	TGCGGCAGTG	GGCGAGCGGG	TTAATCAGAT
	201	ATTTACGTTG	CTGGGANGGG	AAACCGCCTT	GCAAAAGGGG	CAGGCGGGAA
	251	CGGCTCTGGC	AACCTATATG	CTGATGTTGG	AACGCACAAA	ATCCCCGAA
	301	GTGCGCGAAC	GCGCCTTGGG	AATGGCCGTG	TCNCTGAACG	CGTTTGAACA
25	351	GGCGGAAATG	ATTTATCAGA	AATGGCGGCA	GATTGAGCCT	ATACCGGGTA
	401	AGGCGCAAAA	ACGGGCGGGG	TGGCTGCGGA	ACGTGCTGAG	GGAAAGAGGA
	451	AATCAGCATC	TAGACGGACT	GGAGAANTG	CTGGCTCAGG	CGGACGAANG
	501	ACAGAACCGC	AGGGTGTTTT	TATTGTTGGC	ACAAGCCGCC	GTGCAACAGG
	551	ACGGGTTGGC	GCAAAAAGCA	TCGAAAGCGG	TTCGCCGCGC	GGCGTTGAGA
30	601	TATGAACATC	TGCCCCAAGC	GGCGGTTGCC	GATGTGGTGT	TCAGCGTACA
	651	GGNACGCGAA	AAGGAAAAGG	CAATCGGAGC	TTTGCAGCGT	TTGGCGAAGC
	701	TCGATACGGA	AATATTGCCC	CCCACTTTAA	TGACGTTGCG	TCTGACTGCA
	751	CGCAAAATATC	CCGAAATACT	CGACGGCTTT	TTCGAGCAGA	CAGACACCCA
	801	AAACCTTTTCG	GCCGTCTGGC	AGGAAATGGA	AATTATGAAT	CTGGTTTTCCC
35	851	TGCACAGGCT	GGATGATGCC	TATGCGCGTT	TGAACGTGCT	GTTGGAACGC
	901	AATCCGAATG	CAGACCTGTA	TATTCAGGCA	GCGATATTGG	CGGCAAACCG
	951	AAAAGAANGT	GCTTCCGTTA	TCGACGGCTA	CGCCGAAAAG	GCATACGGCA
	1001	GGGGGACGGG	GGAACAGCGG	GGCAGGGCGG	CAATGACGGC	GGCGATGATA
40	1051	TATGCCGACC	GAAGGGATTA	CACCAAAGTC	AGGCAGTGGT	TGAAAAAAGT
	1101	GTCCGCGCCG	GAATACCTGT	TCGACAAAGG	TGTGCTGGCG	GCTGCGGCGG
	1151	CTGTCGAGTT	GGACNGCGGC	AGGGCGGCTT	TGCGGCAGAT	CGGCAGGGTG
	1201	CGGAAACTTC	CCGAACAGCA	GGGGCGGTAT	TTTACGGCAG	ACAATTTGTC
	1251	CAAAATACAG	ATGTTGCCCC	TGTCGAAGCT	GCCCCACAAA	CGGGAGGCTT
45	1301	TGAGGGGGTT	GGACAGGATT	ATCGAAAAAC	CGCCTGCCGG	CAGTAATACA
	1351	GAGTTACAGG	CAGAGGCATT	GGTACAGCGG	TCAGTTGTTT	ACGATCGGCT
	1401	TGGCAAGCGG	AAAAAATGA	TTTCAGATCT	TGAAAGGGCG	TTTCAAGCTTG
	1451	CACCCGATAA	CGCTCAGATT	ATGAATAATC	TGGGCTACAG	CCTGCTTTCC
	1501	GATTCCAAAC	GTTTGGACGA	AGGCTTCGCC	CTGCTTCAGA	CGGCATACCA
50	1551	AATCAACCGG	GACGATACCG	CTGTCAACGA	CAGCATAGGC	TGGGCGTATT
	1601	ACCTGAAAANG	CGACGCGGAA	AGCGCGCTGC	CGTATCTGCG	GTATTCTGTTT
	1651	GAAAACGACC	CCGAGCCCGA	AGTTGCCGCC	CATTTGGGCG	AAGTGTTGTG
	1701	GGCATTGGGC	GAACGCGATC	AGCGGGTTGA	CGTATGGACG	CAGGCGGCAC
	1751	ACCTTACGGG	AGACAAGAAA	ATATGGCGGG	AAACGCTCAA	ACGTACGCGC
	1801	ATCGCATTGC	CCCAACCTTC	CCGAAAACCT	CGGAAATAA	

55 This encodes a protein having amino acid sequence <SEQ ID 44>:

	1	MLPARFTILS	VLAALLAGO	AYAAGAADAK	PPKEVGKVFR	KQORYSEEEI
	51	KNERARLA	GERVNQIFTL	LGXETALQKG	QAGTALATYM	LMMLERTKSPE
	101	VAERALEMAV	SLNAFEQAEM	IYQKWRQIEP	IPGKAQKRAG	WLRNVLRRER
60	151	NQHLDGLEEX	LAQADEXQNR	RVFLLLAQAA	VQDGLAQKA	SKAVRRAALR
	201	YEHLPEAAVA	DVFSVQXRE	KEKAIGALQR	LAKLDTEILP	PTLMTLRLTA
	251	RKYPEILDGF	FEQTDTONLS	AVWQEMEIMN	LVSLHRLDDA	YARLNVLLER
	301	NPNADLYIQA	AILAANRKEX	ASVIDGYAEK	AYGRGTGEQR	GRAAMTAAMI
	351	YADRRDYTKV	RQWLKKVSAP	EYLFDKGVLA	AAAAVELDXG	RAALRQIGRV
	401	RKLPEQQGRY	FTADNLSKIQ	MFALSKLPDK	REALRGLDKI	IEKPPAGSNT
65	451	ELQAEALVQR	SVVYDRLGKR	KKMISDLERA	FRLAPDNAQI	MNNLGYSLLS
	501	DSKRLDEGFA	LLQYAYQINP	DDTAVNDSIG	WAYYLKXDAE	SALPYLRYSF
	551	ENDPEPEVAA	HLGEVLWALG	ERDQAVDVWT	QAAHLTGDKK	IWRETLKRHG



601 IALPQPSRKPK RK\*

ORF9a and ORF9-1 show 95.3% identity in 614 aa overlap:

5	orf9a.pep	10 20 30 40 50	MLPARFTILSVLAAALLAGQAYAAG--AADAKPPKEVGKVFRKQORYSEEEIKNERARLA
	orf9-1	10 20 30 40 50 60	MLPNRFKMLTVLTATLIAGQVSAAGGGAGDMKQPKEVGKVFRKQORYSEEEIKNERARLA
10	orf9a.pep	60 70 80 90 100 110	AVGERVNQIFTLLGXETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
	orf9-1	70 80 90 100 110 120	AVGERVNQIFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
15	orf9a.pep	120 130 140 150 160 170	EMIQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGLEEXLAQADEXQNRVFLLLAQ
	orf9-1	130 140 150 160 170 180	EMIQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGLEEVLAQADEGQNRVFLLLAQ
20	orf9a.pep	180 190 200 210 220 230	AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVFSVQXREKEKAIGALQRLAKLDTEI
	orf9-1	190 200 210 220 230 240	AAVQQDGLAQKASKAVRRAALKYEHLPEAAVADVFSVQGREKEKAIGALQRLAKLDTEI
25	orf9a.pep	240 250 260 270 280 290	LPPTLMTLRLTARKYPEILDGFFEQTDTQNL SAVWQEMEIMNLVSLHRLDDAYARLNVLL
	orf9-1	250 260 270 280 290 300	LPPTLMTLRLTARKYPEILDGFFEQTDTQNL SAVWQEMEIMNLVSLHRLDDAYARLNVLL
30	orf9a.pep	300 310 320 330 340 350	ERNPNADLYIQAAILAANKEXASVIDGYAEKAYGRCTGEQRGRAAMTAAMIYADRRDYT
	orf9-1	310 320 330 340 350 360	ERNPNADLYIQAAILAANKREGASVIDGYAEKAYGRGTEEQRSRAALTAAMMYADRRDYA
35	orf9a.pep	360 370 380 390 400 410	KVRQWLKKVSAPEYLFDKGVLAAAAVELDXGRAALRQIGRVRKLPEQQGRYFTADNLSK
	orf9-1	370 380 390 400 410 420	KVRQWLKKVSAPEYLFDKGVLAAAAVELDGGRAALRQIGRVRKLPEQQGRYFTADNLSK
40	orf9a.pep	420 430 440 450 460 470	IQMFALSKLPDKREALRGLDKIIEKPPAGSNTTELQAEALVQRSVVYDRLGKRKKMISDLE
	orf9-1	430 440 450 460 470 480	IQMLALSKLPDKREALRGLDKIIEKPPAGSNTTELQAEALVQRSVVYDRLGKRKKMISDLE
45	orf9a.pep	480 490 500 510 520 530	RAFRLAPDNAQIMNNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKXD
	orf9-1	490 500 510 520 530 540	RAFRLAPDNAQIMNNLGYSLLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD
50	orf9a.pep	540 550 560 570 580 590	AESALPYLRYSFENDPEPEVAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETILKR
	orf9-1	550 560 570 580 590 600	AESALPYLRYSFENDPEPEVAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETILKR
55	orf9a.pep	600 610	HGIALPQPSRKPRKX
	orf9-1		HGIALPQPSRKPRKX
60			610
65			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF9 shows 82.8% identity over a 163aa overlap with a predicted ORF (ORF9.ng) from *N. gonorrhoeae*:

5	Orf9	RFKMLTVLTATLIAGQVSAAGGGAGDMKQPKVEVGKVFRRKQORYSEEEIKNERAR	54
	orf9ng	MIMLPARFTILSVLAAALLAGQAYAA--GAADVLPKEVGKVLKRRYSEEEIKNERAR	58
10	orf9	LAAVGERVNIQIFTLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE	114
	orf9ng	LAAVGERVNRVFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE	118
	orf9	QAEMIQKWRQIEPIPGKAQKRAQWLRNVLRRGNQHLDGREEVLAQADEGQ	166
	orf9ng	QAEMIQKWRQIEPIGGAQKPAQWLRNVLKEGGNPHLDRLEEVPAQSDYVHQPMIFLLL	178

15 The ORF9ng nucleotide sequence <SEQ ID 45> was predicted to encode a protein having including acid sequence <SEQ ID 46>:

20	1	MIMLPARFTI	LSVLAAALLA	GQAYAAGAAD	VELPKEVGKV	LRKHRRYSEE
	51	EIKNERARLA	AVGERVNRVF	TLLGGETALQ	KGQAGTALAT	YMLMLERTKS
	101	PEVAERALEM	AVSLNAFEQA	EMIQKWRQI	EPIGGEAQKP	AGWLRNVLKE
	151	GGNPHLDRLE	EVPAQSDYVH	QPMIFLLLVQ	AAVQHGGVAQ	KPSKAVRPAA
	201	YNYEVLPEA	GADAVFCVQG	PQYEKAIQSF	PPCGRNPQTE	NIAPPFNELF
	251	RPTARPISPK	LLQRFRTPEP	NLAKPFRPPG	PEMETYQTGF	PRPLTRNNPT

Amino acids 1-28 are a putative leader sequence, and 173-189 are predicted to be a transmembrane domain.

25 Further sequence analysis revealed the complete length ORF9ng DNA sequence <SEQ ID 47>:

30	1	ATGTTACCCG	CCCGTTTCAC	TATTTTATCT	GTCCTCGCAG	CAGCCCTGCT
	51	TGCCGGACAG	GCGTATGCTG	CCGGCGCGGC	GGATGTGGAG	CTGCCGAAGG
	101	AAGTCGGAAA	GGTTTTAAGG	AAACATCGGC	GTTACAGCGA	GGAAGAAATC
	151	AAAAACGAAC	GCGCACGGCT	TGCGGCAGTG	GGCGAACGGG	TCAACAGGGT
	201	GTTTACGCTG	TTGGGCGGTG	AAACGGCTTT	GCAGAAAGGG	CAGGCGGGAA
	251	CGGCTCTGGC	AACCTATATG	CTGATGTTGG	AACGCACAAA	ATCCCCCGAA
35	301	GTCGCCGAAC	GCGCCTTGGA	AATGGCCGTG	TCGCTGAACG	CGTTTGAACA
	351	GGCGGAAATG	ATTTATCAGA	AATGcgggca	gatcgagcct	ataCcggtg
	401	aggcgcaaaa	accgGcgggG	tggctgcgga	acgtattgaa	ggaagggGGA
	451	aaTCAGCATC	TGGAcgggtt	gaaagaggTG	CtggcgcaAT	cggacgatGT
	501	GCAAAAacgc	aggaTATTTT	TGCTGCTGGT	GCAAGCCGCC	GTGCagcagg
	551	gTGGGGTGGC	TCAAAAAGCA	TCGAAAGCGG	TTGCGcgtgc	GGcgttgaAG
40	601	TATGAACATC	TGCCcgaagc	ggcggTTGCC	GATGcggTGT	TCGCGGTACA
	651	GGGACGCGAA	AAGGAAAagg	caaTCGAAGC	TTTGCAGCGT	TTGGCGAAGC
	701	TCGATACGGA	AATATTGCC	CCCACTTTAA	TGACGTTGCG	TCTGACTGCA
	751	CGCAAATATC	CCGAAATACT	CGACGGCTTT	TTCGAGCAGA	CAGACACCCA
	801	AAACCTTTTCG	GCCGCTTGCC	AGGAAATGGA	AATTATGAAT	CTGTTTCCC
	851	TGCGTAAGCC	GGATGATGCC	TATGCGCGTT	TGAACGTGCT	GTGGAACAC
45	901	AACCCGAATG	CAAACCTGTA	TATTCAGGCG	GCGATATTGG	CGGCAAACCG
	951	AAAAGAAGGT	GCGTCCGTTA	TCGACGGCTA	CGCCGAAAAG	GCATACGGCA
	1001	GGGGGACGGG	GGAACAGCGG	GGCagggcgg	cAATgacggc	GGCGATGATA
	1051	TATGCCGACC	GCAGGGATTA	CGCCAAAGTC	AGGCAGTGGT	TGAAAAAAGT
	1101	GTCCGCGCCG	GAATACCTGT	TCGACAAAGG	CGTGCTGGCG	GCTGCGGCGG
	1151	CTGCCGAATT	GGACGGAGGC	CGGGCGGCTT	TGCGGCAGAT	CGGCAGGGTG
50	1201	CGGAAACTTC	CCGAACAGCA	GGGGCGGTAT	TTTACGGCAG	ACAATTTGTC
	1251	CAAAATACAG	ATGCTCGCCC	TGTCGAAGCT	GCCCGACAAA	CGGGAAGCCC
	1301	TGATCGGGCT	GAACAACATC	ATCGCCAAAC	TTTCGGCGGC	GGGAAGCAGC
	1351	GAACCTTTGG	CGGAAGCATT	GGCACAGCGT	TCCATTATTT	ACGaacAGTT
	1401	cggCAAACGG	GGAAAAATGA	TTGCCGACCT	tgaAACcgcg	CTCAAACCTTA
	1451	CGCCCGATAA	TGCACAAATT	ATGAATAATC	TGGGCTACAG	CCTGCTTTCC
55	1501	GATTCCAAAC	GTTTGGACGA	GGGTTTCGCC	CTGCTTCAGA	CGGCATACCA
	1551	AATCAACCCG	GACGATACCG	CCGTTAACGA	CAGCATAGGC	TGGGCGTATT
	1601	ACCTGAAAGG	CGACcggaA	AGCGCGCTGC	CGTATCTGcg	gtattcgttt
	1651	gAAAACGACC	CCGAGCCCGA	AGTTGCCGCC	CATTTGGGCG	AAGTGTGTG

1701 GGCATTGGGC GAACGCGATC AGGCGGTTGA CGTATGGACG CAGGCGGCAC  
 1751 ACCTTAGGGG AGACAAGAAA ATATGGCGGG AGACGCTCAA ACGCTACGGA  
 1801 ATCGCCTTGC CCGAGCCTTC CCGAAAACCC CGGAAATAA

This encodes a protein having amino acid sequence <SEQ ID 48>:

5           1   MLPARFTILS VLAAALLAGO AYAAGAADVE LPKEVGKVLK KHRRYSEEEI  
           51   KNERARLAIV GERVNRVFTL LGGETALQKG QAGTALATYM LMLERTKSPE  
          101   VAERALEMAV SLNAFEQAEM IYQKWRQIEP IPGEAQKPAG WLRNVLKEGG  
          151   NQHLDGLKEV LAQSDDVQKR RIFLLLVQAA VQGGGVAQKA SKAVRRAALK  
 10           201   YEHLPEAAVA DAVFGVQGRE KEKAIEALQR LAKLDTEILP PTLMTLRLTA  
          251   RKYPEILDGF FEQTDTONLS AVWQEMEIMN LVSLRKPDDA YARLNVLEH  
          301   NPNANLYIQA AILAAANRKEG ASVIDGYAEK AYGRGTGEQR GRAAMTAAMI  
          351   YADRRDYAKV RQWLKKSAP EYLFDKGVLA AAAAAELDGG RAALRQIGRV  
          401   RKLPEQQGRY FTADNLSKIQ MLALSKLPDK REALIGLNNI IAKLSAAGST  
          451   EPLAEALAQR SIIYEQFGKR GKMIADLETA LKLTPDNAQI MNNLGYSLLS  
 15           501   DSKRLDEGFA LLQTAYQINP DDTAVNDSIG WAYYLGDAE SALPYLRYSE  
          551   ENDPEPEVAA HLGEVLWALG ERDQAVDVWT QAAHLRGDKK IWRETLKRYG  
          601   IALPEPSRKP RK\*

ORF9ng and ORF9-1 show 88.1% identity in 614 aa overlap:

20	orf9-1.pep	10	20	30	40	50	60
	orf9ng-1	10	20	30	40	50	
25	orf9-1.pep	70	80	90	100	110	120
	orf9ng-1	60	70	80	90	100	110
30	orf9-1.pep	130	140	150	160	170	180
	orf9ng-1	120	130	140	150	160	170
35	orf9-1.pep	190	200	210	220	230	240
	orf9ng-1	180	190	200	210	220	230
40	orf9-1.pep	250	260	270	280	290	300
	orf9ng-1	240	250	260	270	280	290
45	orf9-1.pep	310	320	330	340	350	360
	orf9ng-1	300	310	320	330	340	350
50	orf9-1.pep	370	380	390	400	410	420
	orf9ng-1	360	370	380	390	400	410
55	orf9-1.pep	430	440	450	460	470	480
	orf9ng-1	420	430	440	450	460	470
60		490	500	510	520	530	540

601 IALPQPSRKPK RK\*

ORF9a and ORF9-1 show 95.3% identity in 614 aa overlap:

5	orf9a.pep	10 20 30 40 50	MLPARFTILSVLAAALLAGQAYAAG--AADAKPPKEVGKVFRKQQRYSSEEIKNERARLA
	orf9-1	10 20 30 40 50 60	MLPNRFKMLTVLTATLIAGQVSAAGGGAGDMKQPKVEVGKVFRKQQRYSSEEIKNERARLA
10	orf9a.pep	60 70 80 90 100 110	AVGERVNQIFTLLGXETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
	orf9-1	70 80 90 100 110 120	AVGERVNQIFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
15	orf9a.pep	120 130 140 150 160 170	EMIQKWRQIEPIPGKAQKRAGWLRNVLREGRNQHLGLEEXLAQADEXQNRVFLLLAQ
	orf9-1	130 140 150 160 170 180	EMIQKWRQIEPIPGKAQKRAGWLRNVLREGRNQHLGLEEVLAQADEGQNRVFLLLAQ
20	orf9a.pep	180 190 200 210 220 230	AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVFSVQXREKEKAIGALQRLAKLDEI
	orf9-1	190 200 210 220 230 240	AAVQQDGLAQKASKAVRRAALKYEHLPEAAVADVFSVQGREKEKAIGALQRLAKLDEI
25	orf9a.pep	240 250 260 270 280 290	LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLHRLDDAYARLNVLL
	orf9-1	250 260 270 280 290 300	LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLHRLDDAYARLNVLL
30	orf9a.pep	300 310 320 330 340 350	ERNPNADLYIQAAILAANRKEASVIDGYAEKAYGRGTGEQGRGAAMTAAMIYADRRDYT
	orf9-1	310 320 330 340 350 360	ERNPNADLYIQAAILAANRKEGASVIDGYAEKAYGRGTEEQRSRAALTAAMMYADRRDYA
35	orf9a.pep	360 370 380 390 400 410	KVRQWLKKVSAPEYLFDKGVLAAAAVELDXGRAALRQIGRVRKLPEQQGRYFTADNLSK
	orf9-1	370 380 390 400 410 420	KVRQWLKKVSAPEYLFDKGVLAAAAVELDGGRAALRQIGRVRKLPEQQGRYFTADNLSK
40	orf9a.pep	420 430 440 450 460 470	IQMFALSKLPDKREALRGLDKIIEKPPAGSNTTELQAEALVQRSVVYDRLGKRKKMISDLE
	orf9-1	430 440 450 460 470 480	IQMLALSKLPDKREALRGLDKIIEKPPAGSNTTELQAEALVQRSVVYDRLGKRKKMISDLE
45	orf9a.pep	480 490 500 510 520 530	RAFRLAPDNAQIMNNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKXD
	orf9-1	490 500 510 520 530 540	RAFRLAPDNAQIMNNLGYSLLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD
50	orf9a.pep	540 550 560 570 580 590	AESALPYLRYSFENDPEPEVAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRITLKR
	orf9-1	550 560 570 580 590 600	AESALPYLRYSFENDPEPEVAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRITLKR
55	orf9a.pep	600 610	HGIALPQPSRKPRKX
	orf9-1	610	HGIALPQPSRKPRKX

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF9 shows 82.8% identity over a 163aa overlap with a predicted ORF (ORF9.ng) from *N. gonorrhoeae*:

5	orf9	RFKMLTVLTATLIAGQVSAAGGGAGDMKQPKVEGVKVRKQORYSEEEIKNERAR	54
	orf9ng	MIMLPARFTILSVLAAALLAGQAYAA--GAADVELPKEVGKVLKRHRRYSEEEIKNERAR	58
10	orf9	LAAVGERVNIQIFTLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE	114
	orf9ng	LAAVGERVNRVFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE	118
	orf9	QAEMIQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGREEVLAQADEGQ	166
	orf9ng	QAEMIQKWRQIEPIPGEAQKPAGWLRNVLKEGGNPHLDRLEEVPQSDYVHQPMIFLLL	178

15 The ORF9ng nucleotide sequence <SEQ ID 45> was predicted to encode a protein having including  
acid sequence <SEQ ID 46>:

20

1	MIMLPARFTI	LSVLAAALLA	<u>GQAYAAGAAD</u>	VELPKVEGVK	LRKHRRYSEE
51	EIKNERARLA	AVGERVNRVF	<u>TLLGGETALQ</u>	KGQAGTALAT	YMLMLERTKS
101	PEVAERALEM	AVSLNAFEQA	EMIQKWRQI	EPIPGEAQKP	AGWLRNVLKE
151	GGNPHLDRLE	EVPAQSDYVH	<u>QPMIFLLLLQ</u>	<u>AAVQHGGAAP</u>	KPSKAVRPAA
201	YNYEVLPETA	GADAVFCVQG	PQYEKAIQSF	PPCGRNPQTE	NIAPPFNELF
251	RPTARPISPK	LLQFFRTEP	NLAKPFRPPG	PEMETYQTGF	PRPLTRNNPT

Amino acids 1-28 are a putative leader sequence, and 173-189 are predicted to be a transmembrane domain.

25 Further sequence analysis revealed the complete length ORF9ng DNA sequence <SEQ ID 47>:

	1	ATGTTACCCG	CCCGTTTCAC	TATTTTATCT	GTCCTCGCAG	CAGCCCTGCT
	51	TGCCGGACAG	GCGTATGCTG	CCGGCGCGGC	GGATGTGGAG	CTGCCGAAGG
	101	AAGTCGGA	GGTTTTAAGG	AAACATCGGC	GTTACAGCGA	GGAAGAAATC
30	151	AAAAACGAAC	CGCGACGGCT	TGCGGCAGTG	GGCGAACGGG	TCAACAGGGT
	201	GTTTACGCTG	TTGGGCGGTG	AAACGGCTTT	GCAGAAAGGG	CAGCGCGGAA
	251	CGGCTCTGGC	AACCTATATG	CTGATGTTGG	AACGCACAAA	ATCCCCCGAA
	301	GTCGCCGAAC	GCGCCTTGGA	AATGGCCGTG	TCGCTGAACG	CGTTTGAACA
	351	GGCGGAAATG	ATTTATCAGA	AATGgcggca	gatcgagcct	ataCcggggtg
35	401	agcgcgaaaa	accgGcgggG	tggtcgcgga	acgtattgaa	ggaagggGGA
	451	aaTCAGCATC	TGGACggggt	gaaagaggTG	tgcgcgcaAT	cggacgagtGT
	501	GCAAAAACgc	aggaTATTTT	TGCTGCTGGT	GCAAGCCGCC	GTGCagcagg
	551	gTGGGGTGGC	TCAAAAAGCA	TCGAAAGCGG	TTCGCcgtgc	GGcgttgagAG
	601	TATGAACATC	TGCCcgaagc	ggcggTTGCC	GATGcggTGT	TCGGCGTACA
40	651	GGGACGCGAA	AAGGAAAagg	caaTCGAAGC	TTTGcAGCGT	TTGGCGTAGA
	701	TCGATACGGA	AATATTGCC	CCCACTTTAA	TGACGTTGCG	TCTGACTGCA
	751	CGCAAATATC	CCGAAATACT	CGACGGCTTT	TTCGAGCAGA	CAGACACCCA
	801	AAACCTTTTCG	GCCGCTCTGGC	AGGAAATGGA	AATTATGAAT	CTGTTTCCCC
	851	TGCGTAAGCC	GGATGATGCC	TATGCGCGTT	TGAACGTGCT	GTTGGAACAC
45	901	AACCCGAATG	CAAACTGTA	TATTCAGGGT	CGCATATTGG	CGGCAAAACCG
	951	AAAAGAAGGT	GCGTCCGTTA	TCGACGGCTA	CGCCGAAAAG	GCATACGGCA
	1001	GGGGGACGGG	GGAACAGCGG	GGCagggcgg	cAATGacggc	GGCGATGATA
	1051	TATGCCGACC	GCAGGGATTA	CGCCAAAGTC	AGGCATGTGT	TGAAAAAAGT
	1101	GTCGCGCCCG	GAATACCTGT	TCGACAAAGG	CGTGCTGGCG	GCTGCGCGGG
50	1151	GTCGCCAATT	GGACGAGGCG	CGGGCGGGTT	TCGGCGAGAT	CGGCAGGGTG
	1201	CGGAAACTTC	CCGAACAGCA	GGGGCGGTAT	TTTACGGCAG	ACAAATTGTG
	1251	CAAATACAG	ATGCTCGCCC	TGTCGAAGCT	CCCCGACAAA	CGGGAAGCCC
	1301	TGATCGGGCT	GAACAACATC	ATCGCCAAAC	TTTCGGCGCG	GGGAAGCACG
	1351	GAACTTTTGG	CGGAAGCATT	GGCAGACGCT	TCCATTATTAT	ACGaacAGTT
55	1401	cggCAAAACG	GGAAAAATGA	TTGCCGACCT	tgaAACcgcg	CTCAAACTTA
	1451	CGCCCGATAA	TGCACAAATT	ATGAATAATC	TGGGCTACAG	CCTGCTTTCC
	1501	GATTCCAAAC	GTTTGGACGA	GGGTTTCGCC	CTGCTTCAGA	CGGCATACCA
	1551	AATCAACCCG	GACGATACCG	CCGTTAACGA	CAGCATAGGC	TGGGCGTATT
	1601	ACCTGAAAGG	CGACgcgggA	AGCGCGCTGC	CGTATCTGcg	gtattcgttt
	1651	gAAAAACGACC	CCGAGCCCGA	AGTTGCCGCC	CATTTGGGCG	AAGTGTGTGT

1701 GGCATTGGGC GAACGCGATC AGGCGGTTGA CGTATGGACG CAGGCGGCAC  
 1751 ACCTTAGGGG AGACAAGAAA ATATGGCGGG AGACGCTCAA ACGCTACGGA  
 1801 ATCGCCTTGC CCGAGCCTTC CCGAAAACCC CGGAAATAA

This encodes a protein having amino acid sequence <SEQ ID 48>:

5           1   MLPARFTILS VLAAALLAGO AYAAGAADVE LPKEVGKVLK KHRRYSEEEI  
           51   KNERARLAHV GERVNRVFTL LGGETALQKG QAGTALATYM LMLERTKSPE  
         101   VAERALEMAV SLNAFEQAEM IYQKWRQIEP IPGEAQKPAG WLRNVLKEGG  
         151   NQHLDELKEV LAQSDDVQKR RIFLLLVQAA VQGGGVAQKA SKAVRRAALK  
 10          201   YEHLPEAAVA DAVFGVQGRE KEKAIEALQR LAKLDTEILP PTLMTLRLTA  
         251   RKYPEILDGF FEQTDTONLS AVWQEMEIMN LVSLRKPDPA YARLNVLEH  
         301   NPNANLYIQA AILANRKEG ASVIDGYAEK AYGRGTGEQR GRAAMTAAMI  
         351   YADRRDYAKV RQWLKKVSAP EYLFDKGVLA AAAAAELDGG RAALRQIGRV  
         401   RKLPEQQGRY FTADNLSKIQ MLALSKLPDK REALIGLNNI IAKLSAAGST  
         451   EPLAEALAQR SIIEQFGKR GKMIADLETA LKLTPDNAQI MNNLGYSLLS  
 15          501   DSKRLDEGFA LLQTAYQINP DDTAVNDSIG WAYYLGDAE SALPYLRYSF  
         551   ENDPEPEVAA HLGEVLWALG ERDQAVDVWT QAAHLRGDKK IWRETLKRYG  
         601   IALPEPSRKP RK\*

ORF9ng and ORF9-1 show 88.1% identity in 614 aa overlap:

20	orf9-1.pep	MLPNRFKMLTVLTATLIAGQVSAAGGGAGDMKQPKVEVGKVF	10	20	30	40	50	60
	orf9ng-1	MLPARFTILSVLAAALLAGQAYAAAG--AADVELPKEVGKVLK	10	20	30	40	50	
25	orf9-1.pep	AVGERVNIQIFTLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA	70	80	90	100	110	120
	orf9ng-1	AVGERVNRVFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA	70	80	90	100	110	
30	orf9-1.pep	EMIIYQKWRQIEPIPGKAQKQKAGWLRNVLRERGNQHLDELVEVLAQADEGQNRVFLLLAQ	130	140	150	160	170	180
	orf9ng-1	EMIIYQKWRQIEPIPGEAQKPAGWLRNVLKEGGNQHLDGLKEVLAQSDDVQKRRIFFLLVQ	130	140	150	160	170	
35	orf9-1.pep	AAVQQDGLAQKASKAVRRAALKYEHLEPEAAVADVFSVQGREKEKAIGALQRLAKLDTEI	190	200	210	220	230	240
	orf9ng-1	AAVQQGGVAQKASKAVRRAALKYEHLEPEAAVADAVFGVQGREKEKAIEALQRLAKLDTEI	190	200	210	220	230	
40	orf9-1.pep	LPPTLMTLRLTARKYPEILDGFFEQTDTONLSAVWQEMEIMNLVSLHRLDDAYARLNVLL	250	260	270	280	290	300
	orf9ng-1	LPPTLMTLRLTARKYPEILDGFFEQTDTONLSAVWQEMEIMNLVSLRKPDPAAYARLNVLL	250	260	270	280	290	
45	orf9-1.pep	ERNPNADLYIQAAAILANRKEGASVIDGYAEKAYGRGTGEQRSRAALTAAMMYADRRDYA	310	320	330	340	350	360
	orf9ng-1	EHNPANLYIQAAAILANRKEGASVIDGYAEKAYGRGTGEQGRGAAMTAAMIYADRRDYA	310	320	330	340	350	
50	orf9-1.pep	KVRQWLKKVSAPEYLFDKGVLAATAAVELDGGRAALRQIGRVKRLPEQQGRYFTADNLSK	370	380	390	400	410	420
	orf9ng-1	KVRQWLKKVSAPEYLFDKGVLAATAAVELDGGRAALRQIGRVKRLPEQQGRYFTADNLSK	370	380	390	400	410	
55	orf9-1.pep	IQMLALSKLPDKREALRGLDKIIEKPPAGSNTLQAEALVQSVVYDRLGKRKKMISDLE	430	440	450	460	470	480
	orf9ng-1	IQMLALSKLPDKREALIGLNNIIAKLSAAGSTEPLAEALAQRSIIIEQFGKRGKMIADLE	430	440	450	460	470	
60			490	500	510	520	530	540

```

orf9-1.pep   RAFRLAPDNAQIMNNLGYSLLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD
              |::|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf9ng-1     TALKLTPDNAQIMNNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD
              480      490      500      510      520      530

              550      560      570      580      590      600
orf9-1.pep   AESALPYLRYSFENDPEPEVAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR
              ||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf9ng-1     AESALPYLRYSFENDPEPEVAHLGEVLWALGERDQAVDVWTQAAHLRGDKKIWRETLKR
              540      550      560      570      580      590

              610
orf9-1.pep   HGIALPQPSRKPRKX
              :||||:|||||
orf9ng-1     YGIALPEPSRKPRKX
              600      610

```

In addition, ORF9ng shows significant homology with a hypothetical protein from *P.aeruginosa*:

```

sp|P42810|YHE3_PSEAE HYPOTHETICAL 64.8 KD PROTEIN IN HEMM-HEMA INTERGENIC REGION
(ORF3)
>gi|1072999|pir||S49376 hypothetical protein 3 - Pseudomonas aeruginosa >gi|557259
(X82071) orf3 [Pseudomonas aeruginosa] Length = 576
Score = 128 bits (318), Expect = 1e-28
Identities = 138/587 (23%), Positives = 228/587 (38%), Gaps = 125/587 (21%)

Query: 67  VFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQAEMIQKWR 126
+++LL E A Q+ + AL+ Y++ ++T+ P V+ERA +A L A ++A W
Sbjct: 53  LYSLLVAELAGQRNRFDIALSNYVVQAQKTRDPGVSERAFRIAEYLGADQEALDTSLLWA 112

Query: 127 QIEPIPGEAQKPAG-----WLRNVLKEGGNQHL DGLKEVLAQSDDVQKRRI 172
+ P +AQ+ A ++ VL G+ H D L A++D + +
Sbjct: 113 RSAPDNLDQAQRAAIQLARAGRYEESMVYMEKVLNGQGDTHFDLALSAETDPDTRAGL 172

Query: 173 FXXXXXXXXXXXXXXXXXKASKAVRRAALKYEHLEAAVADAVFGVQGREKEKAIEALQRLA 232
++ KY + + A+ Q ++A+ L+ +
Sbjct: 173 L-----QSF D H L L K K Y P N N G Q L L F G K A L L L Q Q D G R P D E A L T L L E D N S 214

Query: 233 KLDTEILPPTLMTLRLTARK-----YPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLRKP 287
E+ P L + L + K P + G E D + + + + LV +
Sbjct: 215 ASRHEVAPLLRLSRLLQSMKRSDEALPLLKAGIKEHPDDKRVRLAYARL----LVEQNRL 270

Query: 288 DDAYARLNVLLHNP-----ANLYIQAAI----- 312
DDA A L++ P+ A +Y++ +
Sbjct: 271 DDAKAEFAGLVQQFPDDDDDLRFSALVLCLEAQAWDEARIYLEELVERDSHVDAAHFNLG 330

Query: 313 -LAANRKEGASVIDGYAEKAYGRGTGEQGRGAAMTAAMIYADRRDYAKVRQWLKKSVAPE 371
LA +K+ A +D YA+ G G + T ++ A R D A R + P+
Sbjct: 331 RLAEQKDTARALDEYAQ--VGPGNDFLPAQLRQTDVLLKAGRVDEAAQRLDKARSEQPD 388

Query: 372 YLFDKXXXXXXXXXXXXXXXXXQIGRVRKLPEQQGRYFTADNLSKIQLALSKLPDKR 431
Y A L I+ ALS +
Sbjct: 389 Y-----AIQLYLIEAEALSNNDDQOE 408

Query: 432 EALIGLNNIIAKLSAAGSTEPLAEALQRSIIYEQFGKRGKMIADLETALKLTPDNAQIM 491
+A + + + E L L RS++ E+ +M DL + PDNA +
Sbjct: 409 KAWQAIQEGLKQYP-----EDL-NLLYTRSMLEAKRNDLAQMEKDLRFVIAREPDNAMAL 462

Query: 492 NNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD AESALPYLRYFE 551
N LGY+L + R E L+ A++NPDD A+ DS+GW Y +G A YLR + +
Sbjct: 463 NALGYTLADRTTRYGEARELILKAHKLNPDDPAILD SMGWIN YRQGLADAERYLRQALQ 522

Query: 552 NDPEPEVAHLGEVLWALGERDQAVDVWTQAAHLRGDKKIWRETLKR 598
P+ EVAHLGEVLWA G + A +W + + D + R T+KR
Sbjct: 523 RYPDHEVAHLGEVLWAQGRQGDARAIWREYLDKQPDSDVLRRTIKR 569

gi|2983399 (AE000710) hypothetical protein [Aquifex aeolicus] Length = 545
Score = 81.5 bits (198), Expect = 1e-14
Identities = 61/198 (30%), Positives = 98/198 (48%), Gaps = 19/198 (9%)

Query: 408 GRYFTADNL-SKIQMLALSKLPDKREALIGLNNIIAKLSAAGSTEPLAEALAQ----- 459
G Y A L K ++LA PDK+E L + +K + + L +

```

Sbjct: 335 GNYEDAKRLIEKAKVLA----PDKKEILFLEADYYSKTKQYDKALEILKKLEKDYPNDSR 390

Query: 460 ----RSIIYEQFGKRGKMIADLETALKLTPDNAQIMNNGYSLLS--DSKRLDEGFALLQ 513  
 +I+Y+ G L A++L P+N N LGYSL +R++E L++

Sbjct: 391 VYFMEAIVYDNLGDIKNAEKALRKAIELDPENPDYNYLGYSLLLWYGKERVEEAELIK 450

Query: 514 TAYQINPDDTAVNDSIGWAYYLKGAESALPYLRYSF-ENDPEPEVA AHLGEVLWALGER 572  
 A + +P++ A DS+GW YYLKGD E A+ YL + E +P V H+G+VL +G +

Sbjct: 451 KALEKDPENPAYIDSMGWVYYLKGDYERAMQYLLKALREAYDDP VVNEHVG DVLLKMGYK 510

Query: 573 DQAVDVWVTOAAHLRGDKK 590  
 ++A + + +A L + K

Sbjct: 511 EEARNYYERALKLLEEGK 528

- 15 Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 7

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 49>:

20 1 AACCTCTACG CCGGCCCGCA GACCACATCC GTCATCGCAA ACATCGCCGA  
 51 CAACCTGCAA CTGGCCAAAG ACTACGGCAA AGTACACTGG TTCGCCTCCC  
 101 CGCTCTTCTG GCTCCTGAAC CAACTGCACA ACATCATCGG CAACTGGGGC  
 151 TGGGCGATTA TCGTTTAAAC CATCATCGTC AAAGCCGTAC TGTATCCATT  
 201 GACCAACGCC TCTTACCGCT CTATGGCGAA AATGCGTGCC GCCGCACCCA  
 251 AACTGCAAGC CATCAAAGAG AAATACGGCG ACGACCGTAT GCGCGAACAA  
 25 301 CAGGCGATGA TGCAGCTTTA CACAGACGAG AAAATCAACC CGCTGGGCG  
 351 GCTGCCTGCC TATGCTGTTG CAAATCCCCG TCTTCATCGG ATTGTATTGG  
 401 GCATTGTTCT CCTCCGTAGA ATTGCGCCAG GCACCTTGGC TGGGTTGGAT  
 451 TACCGACCTC AGCCGCGCCG ACCCCTACTA CATCCTGCCC ATCATTATGG  
 501 CGGCAACGAT GTTCGCCCAA ACTTATCTGA ACCGCGCCG GACCGACCCG  
 30 551 ATGCagGCGA AAATGATGAA AATCATGCCG TTGGTTTTCT CsGwCrTGT  
 601 CTTCTTCTTC CCTGCCGGks TGGTATTGTA CTGGGTAGTC AACAACCTCC  
 651 TGACCATCGC CCAGCAATGG CACATCAACC GCAGCATCGA AAAACAACGC  
 701 GCCCAAGGCG AAGTCGTTTC CTAA

This corresponds to the amino acid sequence <SEQ ID 50; ORF11>:

35 1 ..NLYAGPQTTS VIANIADNLQ LAKDYGKVHW FASPLFWLLN QLHNIIGNWG  
 51 WAIIVLTIIV KAVLYPLTNA SYRSMKMR A A P K L Q A I K E K Y G D D R M A Q Q  
 101 Q A M M Q L Y T D E K I N P L G G C L P M L L Q I P V F I G L Y W A L F A S V E L R Q A P W L G W I  
 151 T D L S R A D P Y Y I L P I I M A A T M F A Q T Y L N P P P T D P M Q A K M M K I M P L V F S X X F  
 201 F F F P A G X V L Y W V V N N L L T I A Q Q W H I N R S I E K Q R A Q G E V V S \*

- 40 Further sequence analysis revealed the complete DNA sequence <SEQ ID 51>:

1 ATGGATTTTA AAAGACTCAC GCGGTTTTTC GCCATCGCGC TGGTGATTAT  
 51 GATCGGCTGG GAAAAGATGT TCCCCACTCC GAAGCCAGTC CCCGCGCCCC  
 101 AACAGGCAGC ACAACAACAG GCCGTAACCG CTTCCGCCGA AGCCGCGCTC  
 151 GCGCCCGCAA CGCCGATTAC CGTAACGACC GACACGGTTC AAGCCGTCAT  
 45 201 TGATGAAAAA AGCGCGGACC TCGCGCGGCT GACCCTGCTC AAATACAAAG  
 251 CAACCGGCGA CGAAAATAAA CCGTTCATCC TGT TTGGCGA CGGCAAGAA  
 301 TACACCTACG TCGCCCAATC CGAACTTTTG GACGCGCAGG GCAACAACAT  
 351 TCTAAAAGGC ATCGGCTTTA GCGCACCAGG AAAACAGTAC AGCTTGGAAG  
 401 GCGACAAAGT TGAAGTCCGC CTGAGCGCGC CTGAAACACG CCGTCTGAAA  
 50 451 ATCGACAAAG TTTATACTTT CACCAAAGGC AGCTATCTGG TCAACGTCCG  
 501 CTTGACATC GCCAACGGCA GCGGTCAAAC CGCCAACCTG AGCGCGGACT  
 551 ACCGCATCGT CCGGACAC AGCGAACCCG AGGGTCAAG TTACTTTACC  
 601 CACTCTTACG TCGGCCCTGT TGTTTATACC CCTGAAGGCA ACTTCCAAAA  
 651 AGTCAGCTTT TCCGACTTGG ACGACGATGC CAAATCCGGC AAATCCGAGG  
 55 701 CCGAATACAT CCGCAAAACC CCGACCGGCT GGCTCGGCAT GATTGAACAC  
 751 CACTTCATGT CCACCTGGAT TCTCAAACCT AAAGGCAGAC AAAGCGTTTG  
 801 CGCCGCAAGC GAGTGCAACA TCGACATCAA ACGCCGCAAC GACAAGCTGT  
 851 ACAGCACCAG CGTCAGCGTG CCTTTAGCCG CCATCCAAAA CGGCGCGAAA  
 901 GCCGAAGCCT CCATCAACCT CTACGCGGCG CCGCAGACCA CATCCGTCTAT  
 60 951 CGCAAAACATC GCCGACAACC TGCAACTGGC CAAAGACTAC GGCAAAGTAC



1001 ACTGGTTCGC CTCCCCGCTC TTCTGGCTCC TGAACCAACT GCACAACATC  
 1051 ATCGGCAACT GGGGCTGGGC GATTATCGTT TTAACCATCA TCGTCAAAGC  
 1101 CGTACTGTAT CCATTGACCA ACGCCTCTTA CCGCTCTATG GCGAAAAATG  
 1151 GTGCCGCCGC ACCCAAATG CAAGCCATCA AAGAGAAATA CGGCGACGAC  
 5 1201 CGTATGGCGC AACAAACAGG GATGATGCAG CTTTACACAG ACGAGAAAAT  
 1251 CAACCCGCTG GCGCGCTGCC TGCCTATGCT GTTGCAAATC CCCGTCTTCA  
 1301 TCGGATTGTA TTGGGCATTG TTCGCCTCCG TAGAATTGCG CCAGGCACCT  
 1351 TGCTGGGTT GGATTACCGA CCTCAGCCGC GCCGACCCCT ACTACATCCT  
 1401 GCCCATCATT ATGGCGGCAA CGATGTTGCG CCAAACCTAT CTGAACCCGC  
 10 1451 CGCCGACCGA CCCGATGCAG GCGAAAATGA TGAAAATCAT GCCGTTGGTT  
 1501 TTCTCCGTCA GTTCTTCTT CTTCCTGCC GGTCTGGTAT TGTACTGGGT  
 1551 AGTCAACAAC CTCCTGACCA TCGCCAGCA ATGGCACATC AACCGCAGCA  
 1601 TCGAAAAACA ACGCGCCCAA GCGAAGTCG TTTCTCTAA

This corresponds to the amino acid sequence <SEQ ID 52; ORF11-1>:

15 1 MDEFRLTAFF AIALVIMIGW EKMFPKPKV PAPQAAQQQ AVTASAEAL  
 51 APATPITVTT DTQVAVIDEK SGDRLRLTLL KYKATGDENK PFILFGDGKE  
 101 YTYVAQSELL DAQGNILKG IGFSAPKKQY SLEGDKVEVR LSAPETRGLK  
 151 IDKVYFTTKG SYLVNVRFDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT  
 201 HSYVGPVYVY PEGNFQKVSF SLDLDDAKSG KSEAERYRKT PTGWLGMIEH  
 20 251 HFMSTWILQP KGRQSVCAAG ECNIDIKRRN DKLYSTSVSV PLAAIQNGAK  
 301 AEASINLYAG PQTTSVIANI ADNLQAKDY GKVHWFASPL FWLLNQLHNI  
 351 IGNWGWAIIV LTIIVKAVLY PLTNASYRSM AKMRAAPKL QAIKEKYGDD  
 401 RMAQQQAMMQ LYTDEKINPL GGCLPMLLQI PVFIGLYWAL FASVELRQAP  
 451 WLGWITDLR ADPYIILPII MAATMFAQTY LNPPPTDFMQ AKMMKIMPLV  
 25 501 FSVMFFFFPA GLVLYWVVNN LLTIAQQWHI NRSIEKQRAQ GEVVS\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a 60kDa inner-membrane protein (accession P25754) of *Pseudomonas putida*

ORF11 and the 60kDa protein show 58% aa identity in 229 aa overlap (BLASTp).

30 ORF11 2 LYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLTIIVK 61  
 LYAGP+ S + ++ L+L DYG + + A P+FWLL +H+++GNWGW+IIVLT+++K  
 60K 324 LYAGPKIQSKLKLSPGLELTVDYGFLWFIAQPIFWLLQHIHSLGNWGSIIIVLTMLIK 383  
 ORF11 62 AVLYPLTNASYRSMAMRAAPKLQAIKEKYGDDRRXXXXXXXXXXLYTDEKINPLGGCLPM 121  
 + +PL+ ASYRSM+MRA APKL A+KE++GDDR LY EKNPLGGCLP+  
 35 60K 384 GLFFPLSAASYRSMARMRAVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPI 443  
 ORF11 122 LLQIPVFIGLYWALFASVELRQAPWLGWITDLRADPYIILPII MAATMFAQTYLNPPPT 181  
 L+Q+PVF+ LYW L SVE+RQAPW+ WITDLS DP++ILPIIM ATMF Q LNP P  
 40 60K 444 LVQMPVFLALYWVLLESVEMRQAPWILWITDLSIKDPFFILPIIMGATMFIQQRNLPTPP 503  
 ORF11 182 DPMQAKMMKIMPLVXXXXXXXXXPAAGXVLYWVVNNLLTIAQQWHINRSIE 230  
 DPMQAK+MK+MP++ PAG VLYWVVNN L+I+QQW+I R IE  
 60K 504 DPMQAKVMKMPPIIFTFFFLWFPAGLVLYWVVNNCLISISQQWYITRRIE 552

45 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF11 shows 97.9% identity over a 240aa overlap with an ORF (ORF11a) from strain A of *N.*

*meningitidis*:

50 orf11.pep NLYAGPQTTSVIANIADNLQAKDYGKVHW  
 orf11a IKRRNDKLYSTSVSVPLAAIQNGAKSXASINLYAGPQTTSVIANIADNLQAKDYGKVHW  
 280 290 300 310 320 330  
 55 orf11.pep FASPLFWLLNQLHNIIGNWGWAIIVLTIIVKAVLYPLTNASYRSMAMRAAPKLQAIKE  
 orf11a FASPLFWLLNQLHNIIGNWGWAIIVLTIIVKAVLYPLTNASYRSMAMRAAPKLQAIKE  
 340 350 360 370 380 390

-87-

		100	110	120	130	140	150
5	orf11.pep	KYGDDRMAQQQAMMQLYTDEKINPLGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWI					
	orf11a	KYGDDRMAQQQAMMQLYTDEKINPLGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWI					
		400	410	420	430	440	450
10	orf11.pep	TDLRADPYYILPIIIMAATMFAQTYLNPPTDPMQAKMMKIMPLVFSXXFFFFPAGXVLY					
	orf11a	TDLRADPYYILPIIIMAATMFAQTYLNPPTDPMQAKMMKIMPLVXSXXFFXFPAGLVLY					
		460	470	480	490	500	510
15	orf11.pep	WVNNLLTIAQQWHINRSIEKQRAQGEVVVSX					
	orf11a	WVINLLTIAQQWHINRSIEKQRAQGEVVVSX					
		520	530	540			

The complete length ORF11a nucleotide sequence <SEQ ID 53> is:

20	1	ANGGATTTTA	AAAGACTCAC	NGNGTTTTTC	GCCATCGCAC	TGGTGATTAT
	51	GATCGGATNG	NAAANGATGT	TCCCCACTCC	GAAGCCCGTC	CCCGCGCCCC
	101	AACAGACGGC	ACAACAACAG	GCCGTAANCG	CTTCCGCCGA	AGCCGCGCTC
	151	GCGCCCGNAN	CGCCGATTAC	CGTAACGACC	GACACGGTTC	AAGCCGT CAT
	201	TGATGAAAAA	AGCGGCGACC	TGCGCCGGCT	GACCCTGCTC	AAATACAAAG
25	251	CAACCGGCGA	CNAAAATAAA	CCGTTTATCC	TGTTTGCGCA	CGGCAANAA
	301	TACACCTACN	TCGCCANTC	CGAACTTTTG	GACGCGCAGG	GCAACAACAT
	351	TCTAAAAGGC	ATCGGCTTTA	GCGCACCAGG	AAAACAGTAC	AGCTTGGAAG
	401	GCGACAAAGT	TGAAGTCCGC	CTGAGCGCAC	CTGAAACACG	CGGTCTGAAA
	451	ATCGACAAAG	TTTATACTTT	CACCAAAGGC	AGCTATCTGG	TCAACGTCCG
30	501	CTTCGACATC	GCCAACGGCA	GCGGTCAAAC	CGCCAACCTG	AGCGCGGACT
	551	ACCGCATCGT	CCCGGACCAC	AGCGAACCCG	AGGGTCAAGG	CTACTTTTACC
	601	CACTCTTACG	TCGGCCCTGT	TGTTTATACC	CCTGAAGGCA	ACTTCCAAAA
	651	AGTCAGCTTC	TCCGACTTGG	ACGACGATGC	CAANTCCGN	AAATCCGAGG
	701	CCGAATACAT	CCGCAAAACC	CNGACCGGCT	GGCTCGGCAT	GATTGAACAC
35	751	CACTTCATGT	CCACCTGGAT	CCTCCAACCC	AAAGGCGGAC	AAAGCGTTTG
	801	CGCCGCTGGC	GACTGCNGTA	TNGACATCAA	ACGCGCAAC	GACAAGCTGT
	851	ACAGCACCAG	CGTCAGCGTG	CCTTTAGCCG	CTATCCAAAA	CGGTGCGAAA
	901	TCCNAAGCCT	CCATCAACCT	CTACGCCGGC	CCACAGACCA	CATCNGTTAT
	951	CGCAAAACATC	GCCGACAACC	TGCAACTGGN	CAAAGACTAC	GGCAAAGTAC
40	1001	ACTGGTTTCG	CTCCCCCTC	TTTGGCTTT	TGAACCAACT	GCACAACATC
	1051	ATCGGCAACT	GGGCTGGGC	GATTATCGTT	TTAACCATCA	TCGTCAAAGC
	1101	CGTACTGTAT	CCATTGACCA	ACGCCTCTTA	CCGTTTCGATG	GCGAAAATGC
	1151	GTGCCGCCGC	GCCCAAACTG	CAAGCCATCA	AAGAGAAATA	CGGCGACGAC
	1201	CGTATGGCGC	AGCAACAAGC	CATGATGCAG	CTTTACACAG	ACGAGAAAAA
45	1251	CAACCCGCTG	GGCGGCTGCC	TGCCTATGCT	GTTGCAAATC	CCCGTCTTCA
	1301	TCGGATTGTA	TTGGGATTG	TTCGCCTCCG	TAGAATTGCG	CCAGGCACCT
	1351	TGGCTGGGTT	GGATTACCGA	CCTCAGCCGC	GCCGACCCNT	ACTACATCCT
	1401	GCCCATCATT	ATGGCGGCAA	CGATGTTTCG	CCAAACCTAT	CTGAACCCGC
	1451	CGCCGACCGA	CCCGATGCAG	GCGAAAATGA	TGAAAATCAT	GCCTTTGGTT
50	1501	NTNTCNNNNA	NGTCTTCNN	CTTCCCTGCC	GGTCTGGTAT	TGTAAGGGT
	1551	GATCAACAAC	CTCCTGACCA	TCGCCAGCA	ATGGCACATC	AACCGCAGCA
	1601	TCGAAAAACA	ACGCGCCCAA	GGCGAAGTCG	TTTCCTAA	

This encodes a protein having amino acid sequence <SEQ ID 54>:

55	1	XDFKRLTXFF	AIALVIMIGX	XXMFPTPKPV	PAPQOTAQQQ	AVXASAEAL
	51	APXXPITVTT	DTVQAVIDEK	SGDLRRLTLL	KYKATGDKNK	PFILFGDGKX
	101	YTYXAXSELL	DAQNNILKG	IGFSAPKKQY	SLEGDKVEVR	LSAPETRGLK
	151	IDKVYTFTKG	SYLVNVRFDI	ANGSGQTANL	SADYRIVRDH	SEPEGQGYFT
	201	HSYVGPVVYT	PEGNFQKVSF	SDLDDAXSG	KSEAEYIRKT	XTGWLGMIEH
	251	HFMSTWILQP	KGGQSVCAAG	DCXXDIKRRN	DKLYSTSVSV	PLAAIQNGAK
60	301	SXASINLYAG	PQTSVIANI	ADNLQLXKDY	GKVHWFASPL	FWLLNQLHNI
	351	IGNWGWAIIV	LTIIVKAVLY	PLTNASYRSM	AKMRAAAPKL	QAIKEKYGDD
	401	RMAQQQAMMQ	LYTDEKINPL	GGCLPMLLQI	PVFIGLYWAL	FASVELRQAP
	451	WLGWITDLR	ADPYIILPII	MAATMFAQTY	LNPPPTDPMQ	AKMMKIMPLV
	501	XSXXFFXFPA	GLVLYWVINN	LLTIAQQWHI	NRSIEKQRAQ	GEVVS*

ORF11a and ORF11-1 show 95.2% identity in 544 aa overlap:

5	orf11a.pep	XDFKRLTXFFAIALVIMIGXXXMFPTPKPVPAPQQTAAQQQAVXASAEALAPXXPITVTT
	orf11-1	MDFKRLTAFFAIALVIMIGWEKMFPTPKPVPAPQQAQQQAVTASAEALAPATPITVTT
		10 20 30 40 50 60
10	orf11a.pep	DTVQAVIDEKSGDLRRLTLLKYKATGDXNKPFFILFGDGKXYTYXAXSELLDAQGNNILKG
	orf11-1	DTVQAVIDEKSGDLRRLTLLKYKATGDENKPFILFGDGKEYTYVAQSELLDAQGNNILKG
		70 80 90 100 110 120
15	orf11a.pep	IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYTFTKGSYLVNVRFDIANGSGQTANL
	orf11-1	IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYTFTKGSYLVNVRFDIANGSGQTANL
		130 140 150 160 170 180
20	orf11a.pep	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDDAKSGKSEAEYIRKT
	orf11-1	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDDAKSGKSEAEYIRKT
		190 200 210 220 230 240
25	orf11a.pep	XTGWLGMIEHHFMSTWILQPKGGQSVCAAGDCXXDIKRRNDKLYSTSVSVPLAAIQNGAK
	orf11-1	PTGWLGMIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTSVSVPLAAIQNGAK
		250 260 270 280 290 300
30	orf11a.pep	SXASINLYAGPQTTSVIANIADNLQKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV
	orf11-1	AEASINLYAGPQTTSVIANIADNLQKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV
		310 320 330 340 350 360
35	orf11a.pep	LTIIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKEKYGDDRMAQQQAMQLYTDEKINPL
	orf11-1	LTIIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKEKYGDDRMAQQQAMQLYTDEKINPL
		370 380 390 400 410 420
40	orf11a.pep	GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTY
	orf11-1	GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTY
		430 440 450 460 470 480
45	orf11a.pep	LNPPPTDPMQAKMMKIMPLVXSXXFFXFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ
	orf11-1	LNPPPTDPMQAKMMKIMPLVFSVMFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ
		490 500 510 520 530 540
50	orf11a.pep	GEVVSX
	orf11-1	GEVVSX
		55

# 60 Homology with a predicted ORF from *N.gonorrhoeae*

ORF11 shows 96.3% identity over a 240aa overlap with a predicted ORF (ORF11.ng) from *N. gonorrhoeae*:

65	Orf11	NLYAGPQTTSVIANIADNLQKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLT	57
	orf11ng	MAVNLYAGPQTTSVIANIADNLQKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLT	60

5	orf11	IIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINPLGG	117
	orf11ng	IIVKAVLYPLTNASYRSMAKMRAAAPQLQTIKEKYGDDRMAQQQAMMQLFEDEEINPLGG	120
10	orf11	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLRADPYIILPIIMAATMFAQTYLN	177
	orf11ng	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLRADPYIILPIIMAATMFAQTYLN	180
15	orf11	PPPTDPMQAKMMKIMPLVFSXXFFFFPAGXVLYWVVNNLLTIAQQWHINRSIEKQRAQGE	237
	orf11ng	PPPTDPMQAKMMKIMPLVFSVMFFFFPAGLVLYWVVNNLLTIAQQWHINRSIEKQRAQGE	240
20	orf11	VVS 240	
	orf11ng	VVS 243	

An ORF11ng nucleotide sequence <SEQ ID 55> was predicted to encode a protein having amino acid sequence <SEQ ID 56>:

20

1	MAVNLYAGPQ	TTSVIANIAD	NLQLAKDYGK	VHWFASPLFW	LLNQLHNIIG
51	NWGWAIVVLT	IIVKAVLYPL	TNASYRSMAK	MRAAAPQLQT	IKEYGDDRM
101	AQQQAMMLQF	EDEEINPLGG	CLPMLLQIPV	FIGLYWALFA	SVELLRQAPWL
151	GWITDLSRAD	PYYILPIIMA	ATMFAQTYLN	PPPTDFMQAK	<u>SMVKIMPLVFS</u>
201	VMFFFFPAGL	VLYWVVNNLL	TIAQQWHINR	SIEKROAQGE	VVS*

Further sequence analysis revealed the complete gonococcal DNA sequence <SEQ ID 57> to be:

25	1	ATG GATT TTA	AAAGACTCAC	GCGCTTTTTC	GCCATCGCGC	TGGTGATTAT
	51	GATCGGCTGG	GAAAAAATGT	TCCCCACCCC	GAAACCCGTC	CCCGCGCCCC
	101	AACAGGCGGC	ACAAAAACAG	GCAGCAACCG	CTTCCGCCGA	AGCCGCGTTAT
	151	GCGCCCGCAA	CGCCGATTAC	CGTAACGACC	GACACGGTTT	AAGCCGTTAT
30	201	TGATGAAAAA	AGTGGCGACC	TGCGCGGGCT	GACCTTGCTC	AAATACAAAG
	251	CAACCGGCGA	CGAAAAACAA	CCGTTTCGTC	TGTTTGGCGA	CGGCAAAGAA
	301	TACACCTACG	TCGCCCAATC	CGAACTTTTG	GAGCGCGAGG	GCACAACAT
	351	TCTGAAAGGC	ATCGGCTTTA	GCGCACCGAA	AAAAACAGTA	ACCCTCAACG
35	401	GCGACACAGT	CGAAGTCCGC	CTGAGCGCGA	CGGAAACCAA	CGGACTGAAA
	451	ATCGACAAAG	TCTATACCTT	TACCAAAGAC	AGCTATCTGG	TCAACGTCGG
	501	CTTCGACATC	GCCAACGGCA	GCGGTCAAAC	CGCCAACTTG	AGCGCGGACT
	551	ACCGCATCGT	CCGCGACCAC	AGCGAACCCG	AGGGTCAAGG	CTACTTTACC
40	601	CACCTCTTACG	TCGGCCCTGT	TGTTTATACC	CCTGAAGGA	ACTTCCAAAA
	651	AGTCAGCTTC	TCCgacTTgg	acgACGATGC	gaaaTccggc	aaATccgagg
	701	ccgaatacaT	CCGCAAAACC	ccgaccgggt	ggctcggcat	gattgaacac
	751	cacttcatgt	ccacctggat	cctccAAcct	aaaggcggcc	aaaacgtttg
45	801	cgcccagggg	gactgccgta	tcgacattaa	aCgcccgaac	gacaagctgt
	851	acagcgcaag	cgtcagcggt	cctttaaccg	ctatcccaac	ccggggggca
	901	aaaccgaaaa	tggcggTCAA	CCTGTATGCC	GGTCCGCAAA	CCACATCCGT
	951	TATCGCAAAC	ATCGCcgacA	ACCTGCAACT	GGCAAAAGAC	TACGGTAAAG
50	1001	TACACTGGTT	CGCATCGCCG	CTCTTCTGGC	TCCTGAACCA	ACTGCACAAC
	1051	ATTATCGGCA	ACTGGGGCTG	GGCAATCGTC	GTTTTGACCA	TCATCGTCAA
	1101	AGCCGTACTG	TATCCATTGA	CCAACGcttc	ctACCGTTTC	ATGGCGAAAA
	1151	TGCGTGccgc	cgcacCcaaA	CTGCAGACCA	TCAAAAGAAAA	ATAcgGCGAC
55	1201	GACCGTATGG	CGCAACAGCA	AGCGATGATG	CAGCTTTACA	AAgacgAGAA
	1251	AATCAACCCG	CTGGGCGGCT	GTctgcctat	gctgttgCAA	ATCCCCGCTC
	1301	TCATCGGCTT	GTACTGGGCA	TTGTTTCGCT	CCGTAGAATT	GCGCCAGGTA
	1351	CCTTGGCTGG	GACTGGATTAC	CGACCTCAGC	CGCGCCGACC	CCTACTACAT
	1401	CCTGCCCATC	ATTATGGCGG	CAACGATGTT	CGCCCAAACC	TATCTGAACC
	1451	CGCCGCCGAC	CGACCCGATG	CAGGCGAAAA	TGATGAAAAT	CATGCCGTTG
	1501	GTTTTCTCCG	TCATGTTCTT	CTTCTTCCCT	GCCGGTTTGG	TTCTCTACTG
	1551	GGTGTCTCAAC	AACTCTCTGA	CCATCGCCCA	CGAGTGGCAG	ATCAACCGCA
55	1601	GATCGTAAAA	ACAACGCGCC	CAAGCGGAAG	TCGTTTCTTA	A

This encodes a protein having amino acid sequence <SEQ ID 58; ORF1 lng-1>:

1	MDFKRLTAFF	AIALVIMIGW	EKMFP <sup>1</sup> TPKPV	PAPQQA <sup>2</sup> AQKQ	AATASAE <sup>3</sup> AAL
51	APATPITVTT	DTVQAVIDEK	SGDLRRLTLL	KYKATGDENK	PFVLPFGDGKE
101	YTYVAQSELL	DAQGNNILKG	IGFSAPKKQY	TLNGDTVEVR	LSAPETNGLYK
151	IDKVYTFTKD	SYLVNVRFDI	ANGSGQTANL	SADYRIVRDH	SEPEGQGYFT
201	HSYVGPPVYT	PEGNFQKVSF	SDLDDDAKSG	KSEAEYIRKT	PTGWLGMIEH
251	HFMSTWILQP	KGGQNVCAQG	DCRIDIKRRN	DKLYSASVSV	PLTAIPTRGP
301	KPKMAVNLVA	GPOTTSVIAN	IADNLQ <sup>4</sup> LAKD	YGVK <sup>5</sup> VHWFAS	PFLWLLNQLNH

351 IIGNWGWAIV VLTIIIVKAVL YPLTNASYRS MAKMRRAAPK LQTIKEYGD  
 401 DRMAQQQAMM QLYKDEKINP LGGCLPMLLQ IPVFIGLYWA LFASVELRQA  
 451 PWLGWITDLS RADPYIYILPI IMAATMFAQT YLNPPPTDPM QAKMMKIMPL  
 501 VFSVMFFFFFF AGLVLYWVVN NLLTIAQQWH INRSIEKQRA QGEVVS\*

5 ORF11ng-1 and ORF11-1 shown 95.1% identity in 546 aa overlap:

		10	20	30	40	50	60
	orf11ng-1.pep	MDFKRLTAFFAIALVIMIGWEKMFPTPKVPVAPQQAQQAATASAEAAALAPATPITVTT					
	orf11-1	MDFKRLTAFFAIALVIMIGWEKMFPTPKVPVAPQQAQQAATASAEAAALAPATPITVTT					
10		10	20	30	40	50	60
	orf11ng-1.pep	DTVQAVIDEKSGDLRRLTLLKYKATGDENKPFVLFQDGKEYTYVAQSELLDAQGNILKG					
15	orf11-1	DTVQAVIDEKSGDLRRLTLLKYKATGDENKPFVLFQDGKEYTYVAQSELLDAQGNILKG					
		70	80	90	100	110	120
	orf11ng-1.pep	IGFSAPKKQYTLNGDTVEVRLSAPETNGLKIDKVYFTKDSYLVNVRFDIANGSGQTANL					
20	orf11-1	IGFSAPKKQYSLGDKVEVRLSAPETRGLKIDKVYFTKGSYLVNVRFDIANGSGQTANL					
		130	140	150	160	170	180
	orf11ng-1.pep	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDLDDAKSGKSEAEYIRKT					
25	orf11-1	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDLDDAKSGKSEAEYIRKT					
		190	200	210	220	230	240
	orf11ng-1.pep	PTGWLGMIEHHFMSTWILQPKGGQNVCAQGD CRIDIKRRNDKLYSASVSVPLTAIPTRGP					
30	orf11-1	PTGWLGMIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTSVSVPLAAIQN-GA					
		250	260	270	280	290	
35	orf11ng-1.pep	KPKMAVNLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIV					
40	orf11-1	KAEASINLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAII					
		300	310	320	330	340	350
	orf11ng-1.pep	VLTIIIVKAVLYPLTNASYRSMAKMRAAPKLQTIKEYGDDRMAQQQAMMQLYKDEKINP					
45	orf11-1	VLTIIIVKAVLYPLTNASYRSMAKMRAAPKLQAIKEYGDDRMAQQQAMMQLYTDEKINP					
		360	370	380	390	400	410
	orf11ng-1.pep	LGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYIYILPIIMAATMFAQT					
50	orf11-1	LGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYIYILPIIMAATMFAQT					
		420	430	440	450	460	470
	orf11ng-1.pep	YLNPPPTDPMQAKMMKIMPLVFSVMFFFFFFPAGLVLYWVVNNLLTIAQQWHINRSIEKQRA					
55	orf11-1	YLNPPPTDPMQAKMMKIMPLVFSVMFFFFFFPAGLVLYWVVNNLLTIAQQWHINRSIEKQRA					
		480	490	500	510	520	530
60	orf11ng-1.pep	QGEVVSX					
	orf11-1	QGEVVSX					
		540					

65 In addition, ORF11ng-1 shows significant homology with an inner-membrane protein from the database (accession number p25754):

10

15

20

25

30

35

40

45

50

55

60

65

```

      510      520      530      540      550      560
orf11ng-1.pep SVMFFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQGEVVSX
               :|::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
p25754         TFFFLWFEPAGLVLYWVNNCLSLISQQWYITRRIEAATKKAAA
      520      530      540      550      560

```

Based on this analysis, including the homology to an inner-membrane protein from *P. putida* and the predicted transmembrane domains (seen in both the meningococcal and gonococcal proteins), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 5 Example 8

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 59>:

```

1   .GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGGTTGT
51  NAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
101 CGCCTGCCGC CGTCTTGACC GNCGCTCTGC TTCCGCGCT GGGTATTtNG
151 TTCGTACACG CCAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
201 GGATTGGAT GCCGGACAAT ATGTCGAAAT CCTCCGNCAC ACAGGCGGCA
251 ACCGTTACGA AGTT.TTTAT CGCGGTACG. ACTGGCAGGC TCAAAATACG
301 GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTG TCCGCAAGGA
351 AGGCAACCTT CTTATTATCA CACACCCTTA A

```

15 This corresponds to the amino acid sequence <SEQ ID 60; ORF13>:

```

1   .AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
51  FVHAKTAVRK VETDSYQDLG AGQYVEILRH TGGNRYEVXY RGTWQAQNT
101 GQEELEPGTR ALIVRKEGNL LIITHP*

```

Further sequence analysis elaborated the DNA sequence slightly <SEQ ID 61>:

```

20  1   .GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGGTTGT
51  nAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
101 CGCCTGCCGC CGTCTTGACC GnCGCTCTGC TTCCGCGCT GGGTATTtNG
151 TTCGTACACG CCAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
201 GGATTGGAT GCCGGACAAT ATGTCGAAAT CCTCCGACAC ACAGGCGGCA
25  251 ACCGTTACGA AGTTTTtTAT CGCGGTACGc ACTGGCAGGC TCAAAATACG
301 GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTG TCCGCAAGGA
351 AGGCAACCTT CTTATTATCA CACACCCTTA A

```

This corresponds to the amino acid sequence <SEQ ID 62; ORF13-1>:

```

30  1   .AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
51  FVHAKTAVRK VETDSYQDLG AGQYVEILRH TGGNRYEVFY RGTWQAQNT
101 GQEELEPGTR ALIVRKEGNL LIITHP*

```

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF13 shows 92.9% identity over a 126aa overlap with an ORF (ORF13a) from strain A of *N.*

35 *meningitidis*:

```

                                     10      20      30      40      50
orf13.pep      AVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT XALLSALGIXF
|||||
orf13a      MTWVFVA AVAVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT AALLSALGIWF
                                     10      20      30      40      50      60
                                     60      70      80      90     100     110
orf13.pep      VHAKTAVRKVETDSYQDLGAGQYVEILRH TGGNRYEVXYRGTWQAQNTGQEELEPGTRA
|||||
45  orf13a      VHAKTAVGKVETDSYQDLGAGQYAEILRHAGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
                                     70      80      90     100     110     120
                                     120
orf13.pep      LIVRKEGNLLIITHPX
50  |||||

```

orf13a LIVRKEGNLLIIAKPX  
130

The complete length ORF13a nucleotide sequence <SEQ ID 63> is:

```

5      1  ATGACTGTAT GGTTTGTTGC CGCTGTTGCC GTCTTAATCA TCGAATTATT
      51  GACGGGAACG GTTTATCTTT TGGTTGTCAG CGCGGCTTTG GCGGGTTCGG
     101  GCATTGCTTA CGGGCTGACC GGCAGCACGC CTGCCGCCGT CTTGACCGCC
     151  GCTCTGCTTT CCGCGCTGGG TATTTGGTTC GTACACGCCA AAACCGCCGT
     201  GGGAAAAGTT GAAACGGATT CATATCAGGA TTTGGATGCC GGGCAATATG
     251  CCGAAATCCT CCGGCACGCA GGCGGCAACC GTTACGAAGT TTTTATCGC
    10   301  GGTACGCACT GGCAGGCTCA AAATACGGGG CAAGAAGAGC TTGAACCAGG
     351  AACGCGCGCC CTAATCGTCC GCAAGGAAGG CAACCTTCTT ATCATCGCAA
     401  AACCTTAA
  
```

This encodes a protein having amino acid sequence <SEQ ID 64>:

```

15      1  MTVWFVAAVA VLIIELLTGT VYLLVSAAL AGSGIAYGLT GSTPAAVLTA
      51  ALLSALGIWF VHAKTAVGKV ETDSYQDLDA GQYAEILRHA GGNRYEVFYR
     101  GTHWQAQNTG QEELEPGTRA LIVRKEGNLL IIAKP*
  
```

ORF13a and ORF13-1 show 94.4% identity in 126 aa overlap

```

20      10      20      30      40      50      60
    orf13a.pep  MTVWFVAAVAVLIIELLTGTVYLLVVSAAAGSGIAYGLTGSTPAAVLTAALLSALGIWF
    orf13-1      AVLIIELLTGTVYLLVVSAAAGSGIAYGLTGSTPAAVLTXALLSALGIXF
                  10      20      30      40      50

25      70      80      90      100     110     120
    orf13a.pep  VHAKTAVGKVETDSYQDLDAQYAEILRHAGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
    orf13-1      VHAKTAVRKVETDSYQDLDAQYVEILRHTGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
                  60      70      80      90      100     110

30      130
    orf13a.pep  LIVRKEGNLLIIAKPX
    orf13-1      LIVRKEGNLLIITHPX
                  120
  
```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF13 shows 89.7% identity over a 126aa overlap with a predicted ORF (ORF13.ng) from *N. gonorrhoeae*:

```

40      orf13      AVLIIELLTGTVYLLVVSAAAGSGIAYGLTGSTPAAVLTXALLSALGIXF      51
      orf13ng      MTVWFVAAVAVLIIELLTGTVYLLVVSAAAGSGIAYGLTGSTPAAVLTAALLSALGIWF      60

      orf13      VHAKTAVRKVETDSYQDLDAQYVEILRHTGGNRYEVXYRGTXWQAQNTGQEELEPGTRA      111
      orf13ng      VHAKTAVGKVETDSYQDLDTGKYAEILRYTGGNRYEVFYRGTHWQAQNTGQEVFEPGTRA      120

      orf13      LIVRKEGNLLIITHP      126
      orf13ng      LIVRKEGNLLIIANP      135
  
```

50 The complete length ORF13ng nucleotide sequence <SEQ ID 65> is:

```

55      1  ATGACTGTAT GGTTTGTTGC CGCTGTTGCC GTCTTAATCA TCGAATTATT
      51  GACGGGAACG GTTTATCTTT TGGTTGTCAG CGCGGCTTTG GCGGGTTCGG
     101  GCATTGCCTA CGGGCTGACT GGCAGCACGC CTGCCGCCGT CTTGACCGCC
     151  GCACTGCTTT CCGCGCTGGG CATTTGGTTC GTACATGCCA AAACCGCCGT
     201  GGGAAAAGTT GAAACGGATT CATATCAGGA TTTGGATACC GGAAAATATG
     251  CCGAAATCCT CCGATACACA GGCGGCAACC GTTACGAAGT TTTTATCGC
     301  GGTACGCACT GGCAGGCGCA AAATACGGGG CAGGAAGTGT TTGAACCGGG
     351  AACGCGCGCC CTCATCGTCC GCAAAGAAGG TAACCTTCTT ATCATCGCAA
     401  ACCCTTAA
  
```



This encodes a protein having amino acid sequence <SEQ ID 66>:

```

1  MTVWFVAAVA VLIIELLTGT VYLLVVSAA AGSGIAYGLT GSTPAAVLTA
51  ALLSALGIWF VHAKTAVGKV ETDSYQDLDT GKYEILRYT GGNRYEVFYR
101 GTHWQAQNTG QEVFEPGTRA LIVRKEGNLL IIANP*

```

5 ORF13ng shows 91.3% identity in 126 aa overlap with ORF13-1:

```

10 orf13-1.pep      10      20      30      40      50
      AVLIIELLTGT VYLLVVSAA AGSGIAYGLT GSTPAAVLTXALLSALGIXF
orf13ng      10      20      30      40      50      60
      MTVWFVAAVA VLIIELLTGT VYLLVVSAA AGSGIAYGLT GSTPAAVLTAALLSALGIWF

15 orf13-1.pep      60      70      80      90      100     110
      VHAKTAVRKVETDSYQDL DAGQYVEILRHTGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
orf13ng      60      70      80      90      100     110     120
      VHAKTAVGKVETDSYQDL DTGKYAEILRYTGGNRYEVFYRGTHWQAQNTGQEVFEPGTRA

20 orf13-1.pep      120
      LIVRKEGNLLIITHPX
orf13ng      120
      LIVRKEGNLLIIANPX
      130

```

Based on this analysis, including the extensive leader sequence in this protein, it is predicted that ORF13 and ORF13ng are likely to be outer membrane proteins. It is thus predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 9

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 67>:

```

30 1  ATGTWTGATT TCGGTTTGG CGArCTGGTT TTTGTCGGCA TTATCGCCCT
51  GATWgtCCTC GGCCCCGAAC GCsTGCCCGA GGCCGCCCGC AyCGCCGGAC
101 GGcTCATCGG CAGGCTGCAA CGCTTTGTCTG GcAGCGTCAA ACAGGAATTT
151 GACACTCAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
35 201 AGCTGCCGcC GCTCAGGTTC GAGACAGCCT CAAAGAAACC GGTACGGATA
251 TGGAAGGCAA TCTGCACGAC ATTTCGACG GTCTGAAGCC TTGGGAAAAA
301 CTGCCCCAAC AGCGGACACC TGCCGATTTC GGTGTCTGATG AAAACGGCAA
351 TCCGCT.TCC CGATGCGGCA AACACCCTAT CAGACGGCAT TTCCGACGTT
401 ATGCCGTC..

```

This corresponds to the amino acid sequence <SEQ ID 68; ORF2>:

```

40 1  MXDFLGLGLV FVGIIALIVL GPERXPEAAR XAGRLIGRLQ RFVGSVKQEF
51  DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
101 LPEQRTPADF GVDENGNPXS RCGKHPIRRH FRRYAV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 69>:

```

45 1  ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTCGGCA TTATCGCCCT
51  GATTGTCTCTC GGCCCCGAAC GCCTGCCCGA GGCCGCCCGC ACCGCCGGAC
101 GGCTCATCGG CAGGCTGCAA CGCTTTGTCTG GCAGCGTCAA ACAGGAATTT
151 GACACTCAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
201 AGCTGCCGCC GCTCAGGTTC GAGACAGCCT CAAAGAAACC GGTACGGATA
50 251 TGGAAGGCAA TCTGCACGAC ATTTCGACG GTCTGAAGCC TTGGGAAAAA
301 CTGCCCCAAC AGCGGACACC TGCCGATTTC GGTGTCTGATG AAAACGGCAA
351 TCCGCTTCCC GATGCGGCAA ACACCCTATC AGACGGCATT TCCGACGTTA
401 TGCCGTCGCA ACCTTCCTAC GCTTCCGCCG AAACCCCTGG GGACAGCGGG
451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGCGCATG
501 GCGGGAATAC CTGACTGCTT CTGCCGCCGC ACCCGTCGTA CAGACCGTCG

```

551 AAGTCAGCTA TATCGATACT GCTGTTGAAA CGCCTGTTCC GCACACCACT  
601 TCCCTGCGCA AACAGGCAAT AAGCCGCAAA CGCGATT TTC GTCCGAAACA  
651 CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA

This corresponds to the amino acid sequence <SEQ ID 70; ORF2-1>:

5

1	MFDFGLGELV	FVGIIALIVL	GPERLPPEAR	TAGRLIGRLQ	RFVGSVKQEF
51	DTQIELEELR	KAKQEFEEAA	AQVRDSLKET	GTMMEGNLHD	ISDGLKPWEK
101	LPEQRTPADF	GVDEGNPLP	DAANTLSDGI	SDVMPERSY	ASAETLGD SG
151	QTGSTAEPAE	TQDRAWREY	LTASAAAPVV	QTVEVSYIDT	AVETPVPHHT
201	SLRKQALSRK	RDFRPKHRAK	PKLRVRKS*		

10 Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 71 >:

	1	ATGTTTGATT	TCGGTTTGGG	CGAGCTGGTT	TTTGTGCGCA	TTATCGCCCT
	51	GATTGTCTCT	GGCCCCGAAC	GCCTGCCCGA	GGCCGCCCGC	ACCGCCGGAC
	101	GGCTCATCGG	CAGGCTGCAA	CGCTTTGTCT	GCAGCGTCAA	ACAGGAATTT
15	151	GACACGCAA	TGCAACTGGA	AGAACTAAG	AAGGCAAGC	AGGAATTTGA
	201	AGCTGCCGCT	GCTCAGGTT	GAGACAGCCT	CAAAGAAACC	GGTACCGATA
	251	TGGAGGGTAA	TCTGCACGAC	ATTTCCGACG	GTCTGAAGCC	TTGGGAAAAA
	301	CTGCCCGAAC	AGCGCACGCC	TGCTGATTTT	GGTGTCTGAT	AAAACGCGAA
	351	TCCCTTTCCC	GATGCGGCAG	ACACCCTATT	AGACGGCAT	TCCGACGTTA
20	401	TGCCGTCGGA	ACGTTCTCTAC	GCTTCCGCCG	AAACCCCTGG	GGACAGCGGG
	451	CAAACCGGCA	GTACAGCCGA	ACCCGCGGAA	ACCGACCAAG	ACCGTGCATG
	501	GCGGGAATAC	CTGACTGCTT	CTGCCGCCGC	ACCCGTCTGA	CAGACCGTCG
	551	AAGTCAGCTA	TATCGATACC	GCTGTTGAAA	CCCCGTGTTT	GCATACCACT
	601	TCGCTGCGTA	AACAGGCAAT	AAGCCGCAAA	CGCGATTTGC	GTCTAAATC
	651	CCGCGCCAAA	CCTAAATTGC	CGCTCCGATA	ATCATAA	

25 This encodes a protein having amino acid sequence <SEQ ID 72; ORF2a>:

30

1	MDFDGLGELV	FVGIIALIVL	GPRLPEAAR	TAGRLIGRLQ	RFVGSVKQEF
51	DTQIELEELR	KAKQEFEEAA	AQVRDSLKET	GTDMEGNLHD	ISDGLKPFWEK
101	LPEQRTPADF	GVDENGNNFP	DAANTLLDGI	SDVMPSESY	ASAE TLGDSG
151	QTGSAEPAE	TDQDRAWREY	LTASAAAPVV	QTVEVSYIDT	AVETPVPHHT
201	SLRKAISRK	RDLRPKSRAK	PKLRVRKS*		

The originally-identified partial strain B sequence (ORF2) shows 97.5% identity over a 118aa overlap with ORF2a:

		10	20	30	40	50	60
35	orf2.pep	MXDFGLGELVFVGIIALIVLGP	PERXPEAARXAGRLIGRLQRFVGSVKQ	EFD	TQIELEELR		
	orf2a	MFDFGLGELVFVGIIALIVLGP	PERLPEAARTAGRLIGRLQRFVGSVKQ	EFD	TQIELEELR		
		10	20	30	40	50	60
		<hr/>					
		70	80	90	100	110	120
40	orf2.pep	KAKQEF	EAAAAQVRDSLKETGTDM	EGNLHDISDGLKPWEKLPEQRT	PADFGVD	ENG	NPXS
	orf2a	KAKQEF	EAAAAQVRDSLKETGTDM	EGNLHDISDGLKPWEKLPEQRT	PADFGVD	ENG	NPPF
		70	80	90	100	110	120
45		130					
	orf2.pep	RCGKHPIRRHFRRYAV					
	orf2a	DAANTLLDGISDVMPERSYASAETL	GDSGQTGSTAEPAETDQDRAWREYL	TASAA	PPV		
		130	140	150	160	170	180

50 The complete strain B sequence (ORF2-1) and ORF2a show 98.2% identity in 228 aa overlap:

	orf2a.pep	MFD FGLGELVFVGIIALIVLGPERLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR	60
	orf2-1		
	orf2a.pep	MFD FGLGELVFVGIIALIVLGPERLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR	60
55	orf2a.pep	KAKQEF EAAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLP EQRTPADFGVDENGNPFP	120
	orf2-1		
	orf2a.pep	KAKQEF EAAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLP EQRTPADFGVDENGNPFP	120
	orf2-1		
60	orf2a.pep	DAANTLLDGISDVMPSESYASAETLGDSGQTGSTAEP AETDQDRAWREYLTASAAAPVV	180
	orf2-1		

```

    orf2-1      DAANTLSDGISDVMPERSYASAETLGDSGQTGSTAEPAETDQDRAWREYLTASAAAPVV 180
    orf2a.pep   QTVEVSYIDTAVETVPVHTTSLRKQAISRKRDLRPKSRAKPKLRVRKSX 229
    5          orf2-1      QTVEVSYIDTAVETVPVHTTSLRKQAISRKRDLRPKHRAKPKLRVRKSX 229

```

Further work identified a partial DNA sequence <SEQ ID 73> in *N.gonorrhoeae* encoding the following amino acid sequence <SEQ ID 74; ORF2ng>:

```

10          1  MFDFGLGELI  FVGIIALIVL  GPERLPEAAR  TAGRLIGRLQ  RFVGSVKQEL
          51  DTQIELEELR  KVKQAFEAAA  AQVRDSLKET  DTDQNSLHD  ISDGLKPWEK
          101  LPEQRTPADF  GVDEKGNLSL  RYKKHRIRRH  FRRYAV*

```

Further work identified the complete gonococcal gene sequence <SEQ ID 75>:

```

15          1  ATGTTTGATT  TCGGTTTGGG  CGAGCTGATT  TTTGTCTGGCA  TTATCGCCCT
          51  GATTGTCCTT  GGTCCAGAAC  GCCTGCCCCG  AGCCGCCCGC  ACTGCCGGAC
          101  GGCTTATCGG  CAGGCTGCAA  CGCTTTGTAG  GAAGCGTCAA  ACAAGAACTT
          151  GACACTCAAA  TCGAACTGGA  AGAGCTGAGG  AAGGTCAAGC  AGGCATTCTGA
          201  AGCTGCCGCC  GCTCAGGTTC  GAGACAGCCT  CAAAGAAACC  GATACGGATA
          251  TGCAGAACAG  TCTGCACGAC  ATTTCCGACG  GTCTGAAGCC  TTGGGAAAAA
          301  CTGCCCCGAA  AGCGCACGCG  tgccgatttc  gGTGTCGATg  AAAacggcaa
20          351  tccccctccc  gATACGGCAA  ACACCGTATC  AGACGGCATT  TCCGACGTTA
          401  TGCCGTCTGA  ACGTTCGAT  ACTtccgcCG  AAACCTTGG  GGACGACAGG
          451  CAAACCGGCA  GTACAGCCGA  ACCTGCGGAA  ACCGACAAAG  ACCGCGCATG
          501  GCGGGAATAC  CTGactgctt  ctgccgcgcg  acctgtcgta  Cagaggccg
          551  tcgaagtcag  ctaTATCGAT  ACTGCTGTTG  AAacgcctgT  tccgcaCacc
25          601  acttccctgc  gcaAACAGGC  AATAAACCGC  AAACGCGATT  TttgtccgaA
          651  ACACCGCGCc  aaACCGAAat  tgcgcgtcCG  TAAATCATAA

```

This encodes a protein having the amino acid sequence <SEQ ID 76; ORF2ng-1>:

```

30          1  MFDFGLGELI  FVGIIALIVL  GPERLPEAAR  TAGRLIGRLQ  RFVGSVKQEL
          51  DTQIELEELR  KVKQAFEAAA  AQVRDSLKET  DTDQNSLHD  ISDGLKPWEK
          101  LPEQRTPADF  GVDENGNPLP  DTANTVSDGI  SDVMPERSD  TSAETLGDDR
          151  QTGSTAEPAE  TDKDRAWREY  LTASAAAPVV  QRAVEVSYID  TAVETVPVHT
          201  TSLRKQAINR  KRDFCPKHRA  KPCLRVRKS*

```

The originally-identified partial strain B sequence (ORF2) shows 87.5% identity over a 136aa overlap with ORF2ng:

```

35          orf2.pep   MXDFGLGELVFVGIIALIVLGPXPEAARXAGRLIGRLQRFVGSVKQEFDTQIELEELR  60
          orf2ng      MFDFGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR  60
          orf2.pep   KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPXS  120
          orf2ng      KVKQAFEAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDEKGNLSP  120
          orf2.pep   RCGKHPIRRHFRRYAV  136
          orf2ng      RYKKHRIRRHFRYAV  136
45

```

The complete strain B and gonococcal sequences (ORF2-1 & ORF2ng-1) show 91.7% identity in 229 aa overlap:

```

50          orf2-1.pep  MFDFGLGELVFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR
          orf2ng-1     MFDFGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR
          orf2-1.pep  KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPLP
          orf2ng-1     KVKQAFEAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDENGNPLP
          orf2-1.pep  KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPLP
          orf2ng-1     KVKQAFEAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDENGNPLP

```

-97-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf2-1.pep	DAANTLS	SDGISD	VMPSE	RSYASA	ETLGD	SGQTG
	orf2ng-1	DTANTV	SDGISD	VMPSE	RSYASA	ETLGD	SGQTG
		130	140	150	160	170	180
10	orf2-1.pep	Q-TVEV	SYIDTA	VETPV	PHTTSL	RKQAI	SRKR
	orf2ng-1	QRAVE	SYIDTA	VETPV	PHTTSL	RKQAI	SRKR
		190	200	210	220	229	
		Q-TVEV	SYIDTA	VETPV	PHTTSL	RKQAI	SRKR
		190	200	210	220	230	

Computer analysis of these amino acid sequences indicates a transmembrane region (underlined),  
 and also revealed homology (59% identity) between the gonococcal sequence and the TatB protein  
 of *E.coli*:

gnl|PID|e1292181 (AJ005830) TatB protein [Escherichia coli] Length = 171  
 Score = 56.6 bits (134), Expect = 1e-07  
 Identities = 30/88 (34%), Positives = 52/88 (59%), Gaps = 1/88 (1%)  
 Query: 1 MFD FGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60  
 MFD G EL+ V II L+VLGP+RLP A +T I L+ +V+ EL +++L+E +  
 Sbjct: 1 MFDIGFSELLLVFIIGLVVLGPQRLPVAVKTVAGWIRALRSLATTVQNELTQELKLQEFQ 60  
 Query: 61 -KVKQAFEAAAAQVRDSLKETDTDMQNS 87  
 +K+ +A+ + LK + +++ +  
 Sbjct: 61 DSLKKVEKASLTNLTPELKASMDLRQA 88

Based on this analysis, it was predicted that ORF2, ORF2a and ORF2ng are likely to be membrane  
 proteins and so the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be  
 useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF2-1 (16kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above.  
 The products of protein expression and purification were analyzed by SDS-PAGE. Figure 3A  
 shows the results of affinity purification of the GST-fusion protein, and Figure 3B shows the results  
 of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice,  
 whose sera were used for Western blots (Figure 3C), ELISA (positive result), and FACS analysis  
 (Figure 3D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is  
 a useful immunogen.

### Example 10

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 77>:

1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC  
 51 CGC.TGCGGG AACTGACAG GTATTCCATC GCATGGCGgA GkTAAACgCT  
 101 TTgCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA  
 151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC  
 201 CACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGG TCGCTACTCC  
 251 ATTGATGCAC kGrTwCsTGG CGAATACATA AACAGCCCTG CCGTCCGTAC  
 301 CGATTACACC TATCCACGTT ACGAAACCAC CGCTGAAACA ACATCAGGCG  
 351 GTTTGACAGG TTAAACCACT TCTTTATCTA CACTTAATGC CCCTGCACTC  
 401 TCTCGCACCC AATCAGACGG TAGCGGAAGT AAAAGCAGTC TGGGCTTAAA  
 451 TATTGGCGGG ATGGGGGATT ATCGAAATGA AACCTTGACG ACTAACCCGC

-98-

501 GCGACACTGC CTTTCTTTCC CACTTGGTAC AGACCGTATT TTTCCTGCGC  
 551 GGCATAGACG TTGTTTCTCC TGCCAATGCC GATACAGATG TGTTTATTAA  
 601 CATCGAÇGTA TTCGGAACGA TACGCAACAG AACCGAAATG..

This corresponds to the amino acid sequence <SEQ ID 78; ORF15>:

5 1 MQARLLIPIL FSVFILLSACG TLTGIPSHGG XKRFAVEQEL VAASARAANK  
 51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDAXXXG EYINSPAVRT  
 101 DYTYPYRYET AETTSGLTGG LTTSLSLTLNA PALSRQSDG SGSKSSLGLN  
 151 IGGMGDYRNE TLTTNPRDTA FLHLVQTVF FLRGIDVVP ANADTDVFIN  
 201 IDVFGTIRNR TEM..

10 Further work revealed the complete nucleotide sequence <SEQ ID 79>:

1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC  
 51 CGCCTGCGGG AACTGACAG GTATTCCATC GCATGGCGGA GGTAACGCT  
 101 TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA  
 151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC  
 15 CACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA  
 251 TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC  
 301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG  
 351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT  
 401 CTCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT  
 20 451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG  
 501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCTGCGCG  
 551 GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAC  
 601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA  
 651 TGCCGAAACA CTGAAAGCCC AAACAAAAC GGAATATTTT GCAGTAGACA  
 25 701 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT  
 751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA  
 801 AGGAATTAAA CCGACGGAAG GATTAATGGT CGATTCTCTC GATATCCGAC  
 851 CATACGGCAA TCATACGGGT AACTCCGCCC CATCCGTAGA GGCTGATAAC  
 901 AGTCATGAGG GGTATGGATA CAGCGATGAA GTAGTGCGAC AACATAGACA  
 30 951 AGGACAACCT TGA

This corresponds to the amino acid sequence <SEQ ID 80; ORF15-1>:

1 MQARLLIPIL FSVFILLSACG TLTGIPSHGG GKRFAVEQEL VAASARAANK  
 51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT  
 101 DYTYPYRYET AETTSGLTGG LTTSLSLTLNA PALSRQSDG SGSKSSLGLN  
 151 IGGMGDYRNE TLTTNPRDTA FLHLVQTVF FLRGIDVVP ANADTDVFIN  
 201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA  
 251 AYKENYALWM GPYKVSIGIK PTEGLMVDFS DIRPYGNHTG NSAPSVADN  
 301 SHEGYGYSDE VVRQHRQGP \*

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 81>:

40 1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC  
 51 CGCCTGCGGG AACTGACAG GTATTCCATC GCATGGCGGA GGTAACGCT  
 101 TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA  
 151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC  
 45 201 AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA  
 251 TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC  
 301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG  
 351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT  
 401 CGCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT  
 451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG  
 50 501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCTGCGCG  
 551 GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACGGATGT GTTTATTAAC  
 601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA  
 651 TGCCGAAACA CTGAAAGCCC AAACAAAAC GGAATATTTT GCAGTAGACA  
 701 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT  
 55 751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGACCGTATA AAGTAAGCAA  
 801 AGGAATTAAA CCGACAGAAG GATTAATGGT CGATTCTCTC GATATCCAAC  
 851 CATACGGCAA TCATATGGGT AACTCTGCCC CATCCGTAGA GGCTGATAAC  
 901 AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCGAC GACATAGACA  
 951 AGGGCAACCT TGA

60 This encodes a protein having amino acid sequence <SEQ ID 82; ORF15a>:

1 MQARLLIPIL FSVFILLSACG TLTGIPSHGG GKRFAVEQEL VAASARAANK

51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT  
 101 DYTYPYRETT AETTSGLTG LTSLSTLNA PALSRTQSDG SGSSSLGLN  
 151 IGGMGDYRNE TLTNPRDTA FLSHLVQTVF FLRGIDVVS ANADTDVFIN  
 201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA  
 251 AYKENYALWM GPYKVSQGIK PTEGLMVDFS DIQPYGNHMG NSAPSVEADN  
 301 SHEGYGYSDE AVRRHRQGP \*

The originally-identified partial strain B sequence (ORF15) shows 98.1% identity over a 213aa overlap with ORF15a:

10	orf15.pep	10 20 30 40 50 60	MQARLLIPILESVFILSACGTLTGIPSHGGXKRFAVEQELVAASARA AVKMDLQALHGR
	orf15a	10 20 30 40 50 60	MQARLLIPILESVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA AVKMDLQALHGR
15	orf15.pep	70 80 90 100 110 120	KVALYIATMGDQSGSLTGGRYSIDAXXGEYINSPAVRTDYTYPRYETTAETTSGLTG
	orf15a	70 80 90 100 110 120	KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGLTG
20	orf15.pep	130 140 150 160 170 180	LTSLSTLNAPALSRTQSDGSGSSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF
	orf15a	130 140 150 160 170 180	LTSLSTLNAPALSRTQSDGSGSSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF
25	orf15.pep	190 200 210	FLRGIDVVS PANADTDVFINIDVFGTIRNRTEM
	orf15a	190 200 210 220 230 240	FLRGIDVVS PANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL

The complete strain B sequence (ORF15-1) and ORF15a show 98.8% identity in 320 aa overlap:

35	orf15a.pep	10 20 30 40 50 60	MQARLLIPIILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA AVKMDLQALHGR
	orf15-1	10 20 30 40 50 60	MQARLLIPIILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA AVKMDLQALHGR
40	orf15a.pep	70 80 90 100 110 120	KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGLTG
	orf15-1	70 80 90 100 110 120	KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGLTG
45	orf15a.pep	130 140 150 160 170 180	LTSLSTLNAPALSRTQSDGSGSSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF
	orf15-1	130 140 150 160 170 180	LTSLSTLNAPALSRTQSDGSGSSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF
50	orf15a.pep	190 200 210 220 230 240	FLRGIDVVS PANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
	orf15-1	190 200 210 220 230 240	FLRGIDVVS PANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
55	orf15a.pep	250 260 270 280 290 300	IKPKTNAFEAAAYKENYALWMGPYKVSQGIKPTGLMVDFS DIQPYGNHMGNSAPSVEADN
	orf15-1	250 260 270 280 290 300	IKPKTNAFEAAAYKENYALWMGPYKVSQGIKPTGLMVDFS DIRPYGNHTGNSAPSVEADN
60	orf15a.pep	310 320	SHEGYGYSDEAVRRHRQGPX
	orf15-1	310 320	SHEGYGYSDEVVRQHRQGPX

310                      320

Further work identified the corresponding gene in *N.gonorrhoeae* <SEQ ID 83>:

	1	ATGCGGGCAC	GGCTGCTGAT	ACCTATTCTT	TTTTCAGTTT	TTATTTTATC
5	51	CGCCTGCGGG	ACACTGACAG	GTATTCCATC	GCATGGCGGA	GGCAAACGCT
	101	TCGCGGTGCA	ACAAGAACTT	GTGGCCGCTT	CTGCCAGAGC	TGCCGTTAA
	151	GACATGGATT	TACAGGCATT	ACACGGACGA	AAAGTTGCAT	TGTACATTGC
	201	AACATATGGGC	GACCAAGGTT	CAGGCAGTTT	CACAGGGGGT	CGCTACTCCA
	251	TTGATGCACT	GATTTCGCGGC	GAATACATAA	ACAGCCCTGC	CGTCCGCACC
10	301	GATTACACCT	ATCCGCGTTA	CGAAACCACC	GCTGAAACAA	CATCAGGCGG
	351	TTTGACGGGT	TTAACCACTT	CTTTATCTAC	ACTTAATGCC	CCTGCACTCT
	401	CGCGCACCCA	ATCAGACGGT	AGCGGAAGTA	GGCAGAGTCC	GGGCTTAAAT
	451	ATTGGCGGGA	TGGGGGATTA	TCGAAATGAA	ACCTTGACGA	CCAAACCGCG
	501	CGACACTGCC	TTTCTTTCCC	ACTTGGTGCA	GACCGTATTT	TTCCTGCGCG
15	551	GCATAGACGT	TGTTTCTCCT	GCCAATGCCG	ATACAGATGT	GTTTATTAAC
	601	ATCGACGTAT	TCGGAACGAT	ACGCAACAGA	ACCGAAATGC	ACCTATACAA
	651	TGCCGAAACA	CTGAAAGCCC	AAACAAAAC	GGAAATTTTC	CGAGTAGACA
	701	GACCAATAAA	AAAATTGCTC	ATCAAACCCA	AAACCAATGC	GTTTGAAGCT
	751	GCCTATAAAG	AAAATTACGC	ATTGTGGATG	GGGCCGTATA	AAGTAAGCAA
20	801	AGGAATCAAA	CCGACGGAAG	GATTGATGGT	CGATTTCTCC	GATATCCAAC
	851	CATACGGCAA	TCATACGGGT	AACTCCGCCC	CATCCGTAGA	GGCTGATAAC
	901	AGTCATGAGG	GGTATGGATA	CAGCGATGAA	GCAGTGCGAC	AACATAGACA
	951	AGGGCAACCT	TGA			

This encodes a protein having amino acid sequence <SEQ ID 84; ORF15ng>:

25	1	MRARLLIPII	FVSFILSACG	TLTGIPSHGG	GKRFAVEQEL	VAASARAAVK
	51	DMDLQALHGR	KVALYIATMG	DQSGSGLTGG	RYSIDALIRG	EYINSPAVRT
	101	DYTYPRYETT	AETTSGLTGT	LTTSLSTLNA	PALSRTQSDG	SGSRSSLGLN
	151	IGMGDYRNE	TLTTNPRDTA	FLSHLVQTVF	FLRGIDVVSP	ANADTDVFIN
	201	IDVFGTIRN	TEMHLYNAET	LKAQTKLEYF	AVDRTNKKLL	IKPKTNAFEA
30	251	AYKENYALWM	GPYKVSIGIK	PTEGLMVDFF	DIQPYGNHTG	NSAPSVEADN
	301	SHEGYGYSDE	AVRQHRQGP	*		

The originally-identified partial strain B sequence (ORF15) shows 97.2% identity over a 213aa overlap with ORF15ng:

35	orf15.pep	MQARLLIPILFSVFILSACGTLTGIPSHGGGXKRFQELVAASARAQVQKMDLQALHGR	60
	orf15.ng	MRARLLIPILFSVFILSACGTLTGIPSHGGGKRFQELVAASARAQVQKMDLQALHGR	60
40	orf15.pep	KVALYIATMGDQSGSLTGGRYSIDAXXXGEYINSPAVRTDYTPRYETTAETTSGGLTG	120
	orf15.ng	KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGGLTG	120
	orf15.pep	LTTSLSTLNAPALSRTQSDGSGSKSSLGLNIGMGDYRNETLTTNPRDTAFLSHLVQTVF	180
	orf15.ng	LTTSLSTLNAPALSRTQSDGSGSRSSLGLNIGMGDYRNETLTTNPRDTAFLSHLVQTVF	180
45	orf15.pep	FLRGIDVVSANADTDVFINIDVFGTIRNRTEM	213
	orf15.ng	FLRGIDVVSANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL	240

The complete strain B sequence (ORF15-1) and ORF15ng show 98.8% identity in 320 aa overlap:

50	orf15-1.pep	10	20	30	40	50	60
		MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA AVKMDLQALHGR					
	orf15ng	MRARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA AVKMDLQALHGR					
		10	20	30	40	50	60
55	orf15-1.pep	70	80	90	100	110	120
		KVALYIATMGDQGSGLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGGLTG					
	orf15ng	KVALYIATMGDQGSGLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGGLTG					
		70	80	90	100	110	120
60	orf15-1.pep	130	140	150	160	170	180
		LTTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLT TNPRDTAFLSHLVQTVF					

-101-

```

      |||
orff15ng  |||LTTSLSTLNAPALSRTQSDGSGSRSSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF
          130      140      150      160      170      180
5
      190      200      210      220      230      240
orff15-1.pep  FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
      |||
orff15ng  |||FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
          190      200      210      220      230      240
10
      250      260      270      280      290      300
orff15-1.pep  IKPKTNAFEAAYKENYALWMGPYKVSKGIKPTEGLMVD FSDIRPYGNHTGNSAPSVEADN
      |||
orff15ng  |||IKPKTNAFEAAYKENYALWMGPYKVSKGIKPTEGLMVD FSDIQPYGNHTGNSAPSVEADN
          250      260      270      280      290      300
15
      310      320
orff15-1.pep  SHEGYGYSDEVVRQHRQGQPX
      |||
20
orff15ng  |||SHEGYGYSDEAVRQHRQGQPX
          310      320

```

Computer analysis of these amino acid sequences reveals an ILSAC motif (putative membrane lipoprotein lipid attachment site, as predicted by the MOTIFS program).

indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF15-1 (31.7kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 4A shows the results of affinity purification of the GST-fusion protein, and Figure 4B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 4C) and ELISA (positive result). These experiments confirm that ORFX-1 is a surface-exposed protein, and that it is a useful immunogen.

### Example 11

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 85>:

```

35      1  ..GG.CAGCACA AAAAACAGGC GGTGGAACGG AAAAACCGTA TTTACGATGA
      51  TGCCGGGTAT GATATTCGGC GTATTCACGG GCGCATCTC CGCAAAATAT
      101  ATCCCCGCGT TCGGGCTTCA AATTTTCTTC ATCCTGTTT TAACCGCCGT
      151  CGCATTCAAA ACACTGCATA CCGACCTCA GACGGCATCC CGCCCGCTGC
      201  CCGGACTGCC CrGACTGACT GCGGTTTCCA CACTGTTCGG CACAATGTCG
40      251  AGCTGGGTCG GCATAGGCGG CGGTTCACTT TCCGTCCCT TCTTAATCCA
      301  CTGCGGCTTC CCCGCCATA AAGCCATCGG CACATCATCC GGCCTTGCCT
      351  GGCCGATTGC ACTCTCCGGC GCAATATCGT ATCTGCTCAA CGGCCTGAAT
      401  ATTCAGGAT TGCCCGAAGG GTCAC TGGGC TTCCTTTACC TGCCCGCCGT
      451  CGCCGTCCTC AGCGCGGCAA CCATTGCCTT TGCCCGCTC GGTGTCAAAA
45      501  CCGCCACAA ACTTTCTTCT GCCAACTCA AAAAATC.TT CGGCATTATG
      551  TTGCTTTTGA TTGCCGAAA AATGCTGTAC AACCTGCTTT AA

```

This corresponds to the amino acid sequence <SEQ ID 86; ORF17>:

```

      1  ..GQHKKQAVNG KTVFTMMPGM IFGVFTGAFS AKYIPAFGLQ IFFILELTAV
      51  AFKTLHTDPQ TASRPLPLGP XLTA VSTLFG TMSSWVGIGG GSLSVPFLIH

```



-102-

101 CGFPAHKAIG TSSGLAWPIA LSGAISYLLN GLNIAGLPEG SLGFLYLPV  
 151 AVLSAATIAF APLGVKTAHK LSSAKLKSF GIMLLLIAGK MLYNLL\*

Further work revealed the complete nucleotide sequence <SEQ ID 87>:

5 1 ATGTGGCATT GGGACATTAT CTTAATCCTG CTTGCCGTAG GCAGTGGCGG  
 51 AGGTTTATT GCCGGCCTGT TCGGCGTAGG CGGCGGCACG CTGATTGTCC  
 101 CTGTCTGTTT ATGGGTGCTT GATTTCAGG GTTGGCACA ACATCCTTAC  
 151 GCGCAACACC TCGCGTCGG CACATCCTTC GCCGTATGG TCTTCACCGC  
 201 CTTTTCCAGT ATGCTGGGGC AGCACAAAA ACAGGCGGTC GACTGGAAAA  
 251 CCGTATTAC GATGATGCCG GGTATGATAT TCGGCGTATT CACGGGCGCA  
 10 301 CTCTCCGCAA AATATATCCC CGCGTTCGGG CTTCAAATTT TCTTCATCCT  
 351 GTTTTAAACC GCCGTCGCAT TCAAAACACT GCATACCGAC CCTCAGACGG  
 401 CATCCCGCCC GCTGCCGGA CTGCCCGGAC TGAATGCGGT TTCCACACTG  
 451 TTCGGCACAA TGTCGAGCTG GGTCCGCATA GGCGGCGGTT CACTTTCCGT  
 501 CCCCTTCTTA ATCCACTGCG GCTTCCCCGC CCATAAAGCC ATCGGCACAT  
 15 551 CATCCGGCCT TGCCTGGCCG ATTGCACTCT CCGGCGCAAT ATCGTATCTG  
 601 CTCAACGGCC TGAATATTGC AGGATTGCCG GAAGGGTCAC TGGGCTTCCT  
 651 TTACCTGCCG GCCGTGCGCG TCCTCAGCGC GGCAACCATT GCCTTTGCCG  
 701 CGCTCGGTGT CAAAACGCC CACAACTTT CTTCTGCCAA ACTCAAAAA  
 751 Tc.TTCGGCA TTATGTTGCT TTTGATTGCC GGAAAAATGC TGTACAACCT  
 20 801 GCTTTAA

This corresponds to the amino acid sequence <SEQ ID 88; ORF17-1>:

1 MWHWDIILIL LAVGSAAGFI AGLEGVGGGT LIVPVVLWVL DLQGLAQHPY  
 51 AQHLAVGTSF AVMVTAFFSS MLGQHKQAV DWKTVFTMMP GMIFGVFTGA  
 101 LSAKIYPAFG LQIFFILELT AVAFKTLHTD PQTASRPLPG LPGLTAVSTL  
 25 151 FGTMSWVGI GGGSLVFPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL  
 201 LNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKK  
 251 XFGIMLLLIA GKMLYNLL\*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical *H.influenzae* transmembrane protein HI0902 (accession number P44070)

30 ORF17 and HI0902 proteins show 28% aa identity in 192 aa overlap:

ORF17 3 HKKQAVNGKTVFTMMPGMIFGVFT-GAFSAKIYPAFGLQIF--FILFLTAVAFKTLHTDP 59  
 HK + + V + P ++ VF G F + +IF +++L ++ D  
 HI0902 72 HKLGNIVWQAVRILAPVIMLSVFICGLFIGRLDREISAKIFACLVVYLATKMVLSIKKD- 130  
 35 ORF17 60 QTASRPLPGLPXLTA VSTLFGTMSWVGIGGGSLVFPFLIHCGFPAHKAIGTSSGLAWPI 119  
 Q ++ L L + L G SS GIGGG VPFL G +AIG+S+ +  
 HI0902 131 QVTTKSLTFLSSVIG-GILIGMASSAAGIGGGFIVPFLTARGINIKQAIGSSAFCGMLL 189  
 40 ORF17 120 ALSGAISYLLNGLNIAGLPEGSLGFLYLPVAVLSAATIAFAPLGVXXXXXXXXXXXXXXXXX 179  
 +SG S+++G +PE SLG++YLPV ++A + + LG  
 HI0902 190 GISGMFSFIVSGWGNPLMPEYS LGYIYLPV LGITATSFFTSKLGASATAKLPVSTLKKG 249  
 ORF17 180 FGIMLLLIAGKM 191  
 F + L+++A M  
 45 HI0902 250 FALFLIVVAINM 261

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF17 shows 96.9% identity over a 196aa overlap with an ORF (ORF17a) from strain A of *N. meningitidis*:

50 orf17.pep 10 20 30  
 GQHKQAVNGKTVFTMMPGMIFGVFTGAFS  
 orf17a QGLAQHPYAQHLAVGTSFAVMVTAFFSSMLGQHKQAVDWKTVFTMMPGMVFGVFAGALS  
 50 60 70 80 90 100  
 55 orf17.pep 40 50 60 70 80 90  
 AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPXLTA VSTLFGTMSWVGIGG  
 orf17a AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPGLTA VSTLFGTMSWVGIGG

-103-

		110	120	130	140	150	160
		100	110	120	130	140	150
5	orf17.pep	GSLSVPFLLHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPVAV					
	orf17a	GSLSVPFLLHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPVAV					
		170	180	190	200	210	220
10	orf17.pep	AVLSAATIAFAPLGVKTAHKLSSAKLKKSFGIMLLLIAGKMLYNLLX					
	orf17a	AVLSAATIAFAPLGVKTAHKLSSAKLKKSFGIMLLLIAGKMLYNLLX					
		230	240	250	260		

The complete length ORF17a nucleotide sequence <SEQ ID 89> is:

15	1	ATGTGGCATT	GGGACATTAT	CTTAATCCTG	CTTGCCGTAG	GCAGTGCGGC
	51	AGGTTTTTATT	GCCGGCCTGT	TCGGCGTAGG	CGGCGGCACG	CTGATTGTCC
	101	CTGTCGTTTT	ATGGGTGCTT	GATTTGCAGG	GTTTGGCACA	ACATCCTTAC
	151	GCGCAACACC	TCGCCGTCGG	CACATCCTTC	GCCGTCATGG	TCTTCACCGC
	201	CTTTTCCAGT	ATGCTGGGGC	AGCACAAAAA	ACAGGCGGTC	GACTGGAAAA
20	251	CCGTATTTAC	GATGATGCCG	GGTATGGTAT	TCGGCGTATT	CGCTGGCGCA
	301	CTCTCCGCAA	AATATATCCC	AGCGTTCGGG	CTTCAAATTT	TCTTCATCCT
	351	GTTTTTAACC	GCCGTCGCAT	TCAAAACACT	GCATACCGAC	CCTCAGACGG
	401	CATCCCGCCC	GCTGCCCGGA	CTGCCCGGAC	TGACTGCGGT	TTCCACACTG
	451	TTCGGCACAA	TGTCGAGCTG	GGTCGGCATA	GGCGGCGGTT	CACTTTCCGT
25	501	CCCCTTCTTA	ATCCACTGCG	GCTTCCCGCG	CCATAAGCC	ATCGGCACAT
	551	CATCCGGCCT	TGCCTGGCCG	ATTGCACTCT	CCGGCGCAAT	ATCGTATCTG
	601	CTCAACGGCC	TGAATATTGC	AGGATTGCCC	GAAGGTCAC	TGGGCTTCCT
	651	TTACCTGCCC	GCCGTCGCCG	TCCTCAGCGC	GGCAACCATT	GCCTTTGCCC
	701	CGCTCGGTGT	CAAAACCGCC	CACAACTTT	CTTCTGCCAA	ACTCAAAAAA
30	751	TCCTTCGGCA	TTATGTTGCT	TTTGATTGCC	GGAAAAATGC	TGTACAACCT
	801	GCTTTAA				

This encodes a protein having amino acid sequence <SEQ ID 90>:

	1	MWHWDIILIL	LAVGSAAGFI	AGLFGVGGGT	LIVPVVLWVL	DLQGLAQHPY
	51	AQHLAVGTSF	AVMVFTAFSS	MLGQHKQAV	DWKTVFTMMP	GMVFGVFAGA
35	101	LSAKYIPAFG	LQIFFILFLT	AVAFKTLHTD	PQTASRPLPG	LPGLTAVSTL
	151	FGTMSSWVG	GGGSLVFPFL	IHCGFPAHKA	IGTSSGLAWP	IALSGAISYL
	201	LNLNLIAGLP	EGSLGFLYLP	AVAVLSAATI	AFAPLGVKTA	HKLSSAKLKK
	251	SFGIMLLLIA	GKMLYNLL*			

ORF17a and ORF17-1 show 98.9% identity in 268 aa overlap:

40	orf17a.pep	10	20	30	40	50	60
	orf17-1	10	20	30	40	50	60
45	orf17a.pep	70	80	90	100	110	120
	orf17-1	70	80	90	100	110	120
50	orf17a.pep	130	140	150	160	170	180
	orf17-1	130	140	150	160	170	180
55	orf17a.pep	190	200	210	220	230	240
	orf17-1	190	200	210	220	230	240
60	orf17a.pep	250	260	269			
	orf17-1	250	260	269			

```

orf17-1      |||||
              HKLSSAKLKXFGIMLLIAGKMLYNLLX
              250      260

```

5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF17 shows 93.9% identity over a 196aa overlap with a predicted ORF (ORF17.ng) from *N. gonorrhoeae*:

	orf17.pep	GQHKHQAVNGKTVFTMPPGMIFGVFTGAFS	30
		:   :   :           :   :	
10	orf17ng	QGLAQHPYAQH LAVGTSFAVMVFTAFSSMLGQHKHQAVDWKTFAMMPGMIFGVFAGALS	102
	orf17.pep	AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPXLTA VSTLFGTMSSWVGIGG	90
		:	
15	orf17ng	AKYIPAFGLQIFFILFLTAVAFKTLHTGRQTASRPLPGLPGLTAVSTLFGAMSSWVGIGG	162
	orf17.pep	GSLSVPF LIHC GFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAV	150
		:           :	
	orf17ng	GSLSVPF LIHC GFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLPAV	202
20	orf17.pep	AVLSAATIAFAPLG VKTA HKLSSAKLKKSFGIMLLLIAGKM LYNLL	196
		:           :	
	orf17ng	AVLSAATIAFAPLG VKTA HKLSSAKLKESFGIMLLLIAGKM LYNLL	268

An ORF17ng nucleotide sequence <SEQ ID 91> is predicted to encode a protein having amino acid sequence <SEQ ID 92>:

25	1	MWHWDIILIL	LAVGSAAGFI	AGLFGVGGGT	LIVPVVLWVL	DLQGLAQHPY
	51	AQHLAVGTSF	AVMVFTAFSS	MLGQHKKQAV	DWKTIFAMMP	GMIFGVFAGA
	101	LSAKYIPAFG	LQIFFILFLT	AVAFKTLHTG	RQTASRLPG	LPGLTAVSTL
	151	FGAMSSVVGI	GGGSLVFPFL	IHCGFPAHKA	IGTSSGLAWP	IALSGAISYL
	201	VNGLNIAGLP	EGSLGFLYLP	AVAVLSAATI	AFAPLGVKTA	HKLSSAKLKE
30	251	SFGIMLLLIA	GKMLYNLL*			

Further work revealed the complete gonococcal DNA sequence <SEQ ID 93>:

	1	ATGTGGCATT	GGGACATTAT	CTTAATCCTG	CTTGcCgtag	gcAGTGCGGC
	51	AGGTTTTATT	GCCGGCCTGT	Tcgggtgtagg	cggcgGTACG	CTGATTGTCC
35	101	CTGTCTGTTTT	ATGGGTGCTT	GATTTGCAGG	GTTTGGCACA	ACATCCTTAC
	151	GCGCAACACC	TCGCCGTCGG	CAcATccttc	gcCGTCATGG	TCTTCAACCGC
	201	CTTTTCCAGT	ATGTTGGGGC	AGCACAAAA	ACAGGCGGTG	GACTGGAAAA
	251	CCATATTTGC	GATGATGCCG	GGTATGATAT	TCGGCGTATT	CGCTGGCGCA
	301	CTCTCCGCAA	AATATATCCC	CGCGTTCGGG	CTTCAAATTT	TCTTCATCCT
40	351	GTTTTTAACC	GCCGTCGCAT	TCAAAACACT	GCATACCGGT	CGTCAGACGG
	401	CATCCCGCCC	GCTGCCCGGG	CTGCCCGGAC	TGACTGCGGT	TTCCCACTG
	451	TTCGGCGCAA	TGTCGAGCTG	GGTCGGCATA	GGCGGGGGTT	CAC TTCCGT
	501	CCCCTTCTTA	ATCCACTGCG	GCTTCCCGC	CCATAAAGCC	ATCGGCACAT
	551	CATCCGGCCT	TGCCTGGCCG	ATTGCACTCT	CCGGCGCAAT	ATCGTATCTG
	601	GTCAACGGTC	TGAATATTGC	AGGATTGCCC	GAAGGGTCGC	TGGGCTTCCT
45	651	TTACCTGCCC	GCCGTCGCCG	TCCTCAGCGC	GGCAACCAAT	GCCTTTGCCC
	701	CGCTCGGTGT	CAAAACCGCC	CACAACTTT	CTTCTGCCAA	ACTCAAAGAA
	751	TCCTTCGGCA	TTATGTTGCT	TTTGATTGCC	GGAAAAATGC	TGTACAACCT
	801	GCTTTAA				

This corresponds to the amino acid sequence <SEQ ID 94; ORF17ng-1>:

50 1 MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY  
51 AQHLAVGTSF AVMVFTAFSS MLGQHKKQAV DWKTIFAMMP GMIFGVFAGA  
101 LSAKYIPAFG LQIFFILFLT AVAFKTLHTG RQTASRPLPG LPGLTAVSTL  
151 FGAMSSWVGI GGGSLSVFPL IHCGFPAHKA IGTSSGLAWP IALSGAISYL  
201 VNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKE  
55 251 SFGIMLLLIA GKMLYNLL\*

ORF17ng-1 and ORF17-1 show 96.6% identity in 268 aa overlap:

orf17-1.pep MWHWDIILILLAVGSAAGFIAGLFGVGGGTLLIVPVVLWVLDLQGLAQHPYAQHLAVGTSF

-105-

	orfl7ng-1	 MWHWDIILILLAVGSAAGFIAGLFGVGGGTLIVPVVLWVLDLQGLAQHPYAQHLAVGTSF	10	20	30	40	50	60
5	orfl7-1.pep	70 80 90 100 110 120 AVMVFTAFSSMLGQHKKQAVDWKTIFAMMPGMIFGVFTGALS AKYIPAFGLQIFFILFLT						
10	orfl7ng-1	 AVMVFTAFSSMLGQHKKQAVDWKTIFAMMPGMIFGVFTGALS AKYIPAFGLQIFFILFLT	70	80	90	100	110	120
	orfl7-1.pep	130 140 150 160 170 180 AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGGGSLSVPFLIHC GFPAHKA						
15	orfl7ng-1	 AVAFKTLHTGRQTASRPLPGLPGLTAVSTLFGAMSSWVGIGGGSLSVPFLIHC GFPAHKA	130	140	150	160	170	180
	orfl7-1.pep	190 200 210 220 230 240 IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAAVAVLSAATIAFAPLG VKTA						
20	orfl7ng-1	 IGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLPAAVAVLSAATIAFAPLG VKTA	190	200	210	220	230	240
	orfl7-1.pep	250 260 269 HKLSSAKLKKXFGIMLLIAGKMLYNLLX						
25	orfl7ng-1	 HKLSSAKLKESFGIMLLIAGKMLYNLLX	250	260				

In addition, ORF17ng-1 shows significant homology with a hypothetical *H. influenzae* protein:

30	sp P44070 Y902_HAEIN HYPOTHETICAL PROTEIN HI0902 pir G64015 hypothetical protein HI0902 - Haemophilus influenzae (strain Rd KW20) gi 1573922 (U32772) H. influenzae predicted coding region HI0902 [Haemophilus influenzae]Length = 264 Score = 74 (34.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23 Identities = 15/43 (34%), Positives = 23/43 (53%)	
35	Query: 55 AVGTSFAVMVFTAFSSMLGQHKKQAVDWKTIFAMMPGMIFGVF 97 A+GTSFA +V T S HK + W+ + + P ++ VF Sbjct: 52 ALGTSFATIVITGIGSAQRHHKLGNIWQAVRILAPVIMLSVF 94	
40	Score = 195 (91.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23 Identities = 44/114 (38%), Positives = 65/114 (57%)	
45	Query: 150 LFGAMSSWVGIGGGSLSVPFLIHC GFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGL 209 L G SS GIGGG VPFL G +AIG+S+ + +SG S++V+G + Sbjct: 148 LIGMASSAAGIGGGGFIVPFLTARGINIKQAIGSSAFCGMLLGISGMFSFIVSGWGNPLM 207	
50	Query: 210 PEGSLGFLYLPAAVAVLSAATIAFAPLG VKTAHKLSSAKLKESFGIMLLIAGKM 263 PE SLG++YLPVAV ++A + + LG KL + LK+ F + L+++A M Sbjct: 208 PEYSLGYIYLPVAVLGITATSFFTSKLGASATAKLPVSTLKKGFALFLIVVAINM 261	

This analysis, including the homology with the hypothetical *H. influenzae* transmembrane protein, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 12

55 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 95>:

	1	..GGAAACGGAT GGCAGGCAGA CCCC GAACAT CCGCTGCTCG GGCTTTTTCG
	51	CGTCAGTAAT GTATCGATGA CGCTTGCTTT TGTCCGAATA TGTGCGTTGG
	101	TGCATTATTG CTTTTCGGGA ACGGTTCAAG TGTTCGTGTT TGCGGCACTG
	151	CTCAAACTTT ATGCGCTGAA GCCGGTTTAT TGGTTCGTGT TGCAGTTTGT
60	201	GCTGATGGCG GTTGCTATG TCCACCGCTG CCGTATAGAC CGGCAGCCGC
	251	CGTCAACGTT CGGCGGCTCG CAGCTGCGAC TCGGCGGGTT GACGCGAGCG

301 TTGATGCAGG TCTCGGTACT GGTGCTGCTG CTTTCAGAAA TTGGAAGATA  
351 A

This corresponds to the amino acid sequence <SEQ ID 96; ORF18>:

5 1 ..GNGWQADPEH PLLGLFAVSN VSMTLAFVGI CALVHYCFSG TVQVFVFAAL  
51 LKLYALKPVY WFLVQFVLMA VAYVHRCGID RQPPSTFGGS QLRLLGLTAA  
101 LMOVSVLVLL LSEIGR\*

Further work revealed the complete nucleotide sequence <SEQ ID 97>:

	1	ATGATTTTGC	TGCATTTGGA	TTTTTTGTCT	GCCTTACTGT	ATGCGGCGGT
10	51	TTTTCTGTTT	CTGATATCC	GCGCAGGAAT	GTTGCAATGG	TTTTGGGCGA
	101	GTATTATGCT	GTGGCTGGGC	ATATCGGTTT	TGGGGGGCAA	GCTGATGCCC
	151	GGCATATGGG	GAATGACCCG	CGCCGCGCCC	TGTGTCATCC	CCCATTTTTA
	201	CCTGACTTTG	GCGAGCATAT	TTTTTTTCAT	CGGGCATTGG	AACCGGAAAA
	251	CAGATGGAAA	CGGATGGCAG	GCAGACCCCG	AACATCCGCT	GCTCGGGCTT
15	301	TTTGCCGTC	GTAATGTATC	GATGACGCTT	GCTTTTGTGC	GAATATGTGC
	351	GTTGGTGCAT	TATTGCTTTT	CGGGAACGGT	TCAAGTGTTT	GTGTTTGCGG
	401	CACTGCTCAA	ACTTTATGCG	CTGAAGCCGG	TTTATTGGTT	CGTGTTCGAG
	451	TTTGTGCTGA	TGGCGGTTGC	CTATGTCCAC	CGCTGCGGTA	TAGACCGGCA
	501	GCCGCCGTC	ACGTTCCGGC	GCTCGCAGCT	GCGACTCGGC	GGGTTGACGG
20	551	CAGCGTTGAT	GCAGGTCTCG	GTA	TGCTGCTTTC	AGAAATTGGA
	601	AGATAA				

This corresponds to the amino acid sequence <SEQ ID 98; ORF18-1>:

25

1	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQW	FWASIMLWLG	ISVLGAKLMP
51	GIWGMTRAAP	LFIPHFYLT	GSIFFFIGHW	NRKTDGNGWQ	ADPEHPLLGL
101	FAVSNVSMTL	AFVGICALVH	YCFSGTVQVF	VFAALLKLYA	LKPVYWFVLQ
151	FVLMAVAYVH	RCGIDRQPPS	TFGGSQRLRG	GLTAALMQVS	LVLLLLSEIG
201	R*				

Computer analysis of this amino acid sequence gave the following results:

**Homology with a predicted ORF from *N.meningitidis* (strain A)**

ORF18 shows 98.3% identity over a 116aa overlap with an ORF (ORF18a) from strain A of *N.*

30 *meningitidis:*

```

35      orf18.pep      GNGWQADPEHPLLGLFAVSNVSMTLAFVGI
                        |||
      orf18a      TRAAPLFIPHFYLTLSIFFFIHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGI
                    60      70      80      90      100      110

40      orf18.pep      CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGS
                        |||
      orf18a      CALVHYCFSTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGS
                    120      130      140      150      160      170

45      orf18.pep      QLRLGGLTAALMQVSVLVLLLSEIGRX
                        |||
      orf18a      QLRLGGLTAALMQXSVLVLLLSEIGRX
                    180      190      200

```

The complete length ORF18a nucleotide sequence <SEQ ID 99> is:

50	1	ATGATTTTGC	TGCATTTTGA	TTTTTTGTCT	GCCTTACTGT	ATGCGGCGGT
	51	TTTTCTGTTT	CTGATATTCC	GCGCAGGAAT	GTTGCAATGG	TTTTGGGCGA
	101	GTATTATGCT	GTGGCTGGGC	ATATCGGTTT	TGGGGGCAAA	GCTGATGCCC
	151	GGCATATGGG	GAATGACCCG	CGCCGCGCCC	TTGTTTCATC	CCCATTTTAA
	201	CCTGACTTTG	GGCAGCATAT	TTTTTTTTCAT	CGGGCATTTGG	AACCGGAAAA
55	251	CGGATGGAAA	CGGATGGCAG	GCAGACCCCG	AACATCCTCT	GCTCGGGCTG
	301	TTTGCCGTC	GTAATGTATC	GATGACGCTT	GCTTTTGTCT	GAATATGTGC
	351	GTTGGTGTC	TATTGCTTTT	CGNGAACGGT	TCAAGTGTCT	GTGTTTTCGG
	401	CCTGCTGCTA	ACTTTATGCG	CTGAAGCCGG	TTTATTGGTT	CGTGTTCGAG

451 TTTGTGCTGA TGGCGGTTGC CTATGTCCAC CGCTGCGGTA TAGACCGGCA  
 501 GCCGCCGTCA ACGTTCGGCG GNTCGCAGCT GCGACTCGGC GGGTTGACGG  
 551 CAGCGTTGAT GCAGNTCTCG GTACTGGTGC TGCTGCTTTC AGAAATTGGA  
 601 AGATAA

5 This encodes a protein having amino acid sequence <SEQ ID 100>:

1 MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIMLWLG ISVLGAKLMP  
 51 GIWGMTRAAP LFIPHFYLT LGSIFFFIGHW NRKTDGNGWQ ADPEHPLLGL  
 101 FAVSNVSM TLAFVGICALVH YCFSTVQVF VFAALLKLYA LKPVYWFVLQ  
 151 FVLMAYAVH RCGIDRQPPS TFGGSQLRLG GLTAALMQXS VLVLLSEIG  
 201 R\*

ORF18a and ORF18-1 show 99.0% identity in 201 aa overlap:

		10	20	30	40	50	60
	orf18a.pep	MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP					
15	orf18-1	MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP					
		10	20	30	40	50	60
		70	80	90	100	110	120
20	orf18a.pep	LFIPHFYLT LGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSM TLAFVGICALVH					
	orf18-1	LFIPHFYLT LGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSM TLAFVGICALVH					
		70	80	90	100	110	120
25	orf18a.pep	YCFSTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGSQLRLG					
	orf18-1	YCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGSQLRLG					
		130	140	150	160	170	180
30	orf18a.pep	GLTAALMQXS VLVLLSEIGRX					
	orf18-1	GLTAALMQVS VLVLLSEIGRX					
		190	200				
35							

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF18 shows 93.1% identity over a 116aa overlap with a predicted ORF (ORF18.ng) from *N. gonorrhoeae*:

40	orf18.pep	GNGWQADPEHPLLGLFAVSNVSM TLAFVGI	30
	orf18ng	TRAAPLFIPHFYLT LGSIFFFIGYWNKTDGNGWQADPEHPLLGLFAVSNVSM TLAFVGI	115
	orf18.pep	CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGS	90
45	orf18ng	CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGS	175
	orf18.pep	QLRLGGLTAALMQVS VLVLLSEIGR	116
	orf18ng	QLRLGVLAAMLQVAVTAMLLAEIGR	201

50 The complete length ORF18ng nucleotide sequence is <SEQ ID 101>:

1 ATGATTTTGC TGCATTTGGA TTTTGTGCT GCCTTACTGt aTGCGGcggt  
 51 tttTctgTTT CTGATATTCC GCGCAGGAAT GTTGCAATGG TTTTGGGCGA  
 101 GTATTGCGTT GTGGCTCGGC ATCTCGGTTT TAGGGGTAAA GCTGATGCCG  
 55 151 GGGATGTGGG GAATGACCCG CGCCGCGCCT TTGTTCATCC CCCATTTTAA  
 201 CCTGACTTTG GGCAGCATAT TTTTTCAT CCGGTATTGG AACCGGAAAA  
 251 CAGATGGAAA CGGATGGCAG GCAGACCCG AACATCCGCT GCTCGGGCTT  
 301 TTTGCCGTCA GTAATGTATC GATGACGCTT GCTTTGTGCG GAATATGTGC  
 351 GTTGGTGCA TATTGCTTTT CGGGAACGGT TCAAGTGTTT GTGTTTGGCG  
 401 CATTGCTCAA ACTTTATGCG CTGAAGCCGG TTTATTGGTT CGTGTGCGAG  
 60 451 TTTGTATTGA TGGCGGttgC CTATGTCCAC CGCTGCGGTA TAGACCGGCA  
 501 GCCGCCGTCA ACGTTCGGCG GTTCGCAGCT GCGACTCGGC GTGTTGGCGG

551 CGATGTTGAT GCAGGTTGCG GTAACGGCGA TGCTGCTTGC CGAAATCGGC  
601 AGATGA

This encodes a protein having amino acid sequence <SEQ ID 102>:

5 1 MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIALWLGI SVLGVKLMP  
51 GMWGMTRAAP LFIPHFYLTLS GSIFFFIGYW NRKTDGNGWQ ADPEHPLLGL  
101 FAVSNVSM TLAFV GICALVH YCFSGTVQVF VFAALLKLYA LKPVYWFVLQ  
151 FVLMAYAYVH RCGIDRQPPS TFGGSQLRLG VLAAMLMOVA VTAMLLAEIG  
201 R\*

This ORF18ng protein sequence shows 94.0% identity in 201 aa overlap with ORF18-1:

10		10	20	30	40	50	60
	orf18-1.pep	MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP					
	orf18ng						
15		10	20	30	40	50	60
	orf18-1.pep	LFIPHFYLTLSGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSM TLAFV GICALVH					
	orf18ng						
20		70	80	90	100	110	120
	orf18-1.pep	LFIPHFYLTLSGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSM TLAFV GICALVH					
	orf18ng						
25		130	140	150	160	170	180
	orf18-1.pep	YCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAYVHRCGIDRQPPSTFGGSQLRLG					
	orf18ng						
30		190	200				
	orf18-1.pep	GLTAALMQVSVLVLLSEIGRX					
	orf18ng						
		190	200				

Based on this analysis, including the presence of several putative transmembrane domains in the  
35 gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and  
their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 13

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 103>:

40 1 ATGAAAACCC CACTCCTCAA GCCTCTGCTN ATTACCTCGC TTCCCGTTTT  
51 CGCCAGTGTT TTTACCGCCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA  
101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CCTTGTCGAT  
151 TTGGACAACC NCNTGACCGG ACGGCTNAAA AACATCATCA CCACCGTCGC  
201 CCTGTTCACC CTCTCCTCGC TCACGGCACA AAGCACCTC GGCACAGGGC  
251 TGCCCTTCAT CCTCGCCATG ACCCTGATGA CTT.CG.CTT CACCATTTTA  
45 301 GCGCGGNGC ...

This corresponds to the amino acid sequence <SEQ ID 104; ORF19>:

1 MKTPLLKPLL ITS LPVFASV FTAASIVWQL GEPLAMPFV LGIIAGGLVD  
51 LDNXXTGRLK NIITVALFT LSSSLAQSTL GTGLPFILAM TLMTXXFTIL  
101 GAX...

50 Further work revealed the complete nucleotide sequence <SEQ ID 105>:

1 ATGAAAACCC CACTCCTCAA GCCTCTGCTC ATTACCTCGC TTCCCGTTTT  
51 CGCCAGTGTT TTTACCGCCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA  
101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CCTTGTCGAT  
151 TTGGACAACC GCCTGACCGG ACGGCTGAAA AACATCATCA CCACCGTCGC

201 CCTGTTACACC CTCTCCTCGC TCACGGCACA AAGCACCCCTC GGCACAGGGC  
 251 TGCCCTTCAT CCTCGCCATG ACCCTGATGA CCTTCGGCTT CACCATTTTA  
 301 GGCGCGGTCG GGCTCAAATA CCGCACCTTC GCCTTCGGTG CACTCGCCGT  
 351 CGCCACCTAC ACCACACTTA CCTACACCCC CGAAACCTAC TGGCTGACCA  
 401 ACCCCTTCAT GATTTTATGC GGACCGGTAC TGTACAGCAC CGCCATCCTC  
 451 CTGTTCCAAA TCGTCTGCC CCACCGCCCC GTCCAAGAAA GCGTCGCCAA  
 501 CGCCTACGAC GCACTCGGCG GCTACCTCGA AGCCAAAGCC GACTTCTTCG  
 551 ACCCGGATGA GGCAGCCTGG ATAGGCAACC GCCACATCGA CCTCGCCATG  
 601 AGCAACACCG GCGTCATCAC CGCCTTCAAC CAATGCCGTT CCGCCCTGTT  
 651 TTACCGCCTT CGCGGCAAAC ACCGCCACCC GCGCACCGCC AAAATGCTGC  
 701 GTTACTACTT TGCCGCCCAA GACATACACG AACGCATCAG CTCCGCCAC  
 751 GTCGATTATC AGGAAATGTC CGAAAAATTC AAAAACACCG ACATCATCTT  
 801 CCGCATCCAC CGCTGCTCG AAATGCAGGG ACAAGCCTGC CGCAACACCG  
 851 CCCAAGCCCT GCGCGCAAGC AAAGACTACG TTTACAGCAA ACGCCTCGGC  
 901 CGCGCCATCG AAGGCTGCGG CCAATCGCTG CGCCTCCTTT CAGACAGCAA  
 951 CGACAGTCCC GACATCCGCC ACCTGCGCCG CTTTCTCGAC AACCTCGGCA  
 1001 GCGTCGACCA GCAGTTCCGC CAACTCCAGC ACAACGGCCT GCAGGCAGAA  
 1051 AACGACCGCA TGGGCGACAC CCGCATCGCC GCCCTCGAAA CCAGCAGCCT  
 1101 CAAAAACACC TGGCAGGCAA TCCGTCCGCA GCTAAACCTC GAATCAGGCG  
 1151 TATTCGCCA CTGTCCTCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC  
 1201 ATCGTCGAAG CCCTCAACCT CAACCTCGGC TACTGGATAC TACTGACCGC  
 1251 CTTTTTCGTC TGCCAAACCA ACTACACCGC CACCAAAAGC CGCGTCCGCC  
 1301 AGCGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC  
 1351 TACTTCACCC CGTCTGTCTG AACCAACTC TGGATTGTCA TCGCCAGTAC  
 1401 CACCCTCTTT TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT  
 1451 TCATTACCAT TCAAGCCTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA  
 1501 TACGCCGCCA TGCCCGTACG CATCATCGAC ACCATTATCG GCGCATCCCT  
 1551 TGCCTGGGCG GCAGTCAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC  
 1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAACGGTGC CTATCTCGAA  
 1651 AAAATCACCG AACGCCTCAA AAGCGGCGAA ACCGGCGACG ACGTCGAATA  
 1701 CCGCGCCACC CGCCGCCGCG CCCACGAACA CACCGCCGCC CTCAGACGCA  
 1751 CCCTTTCCGA CATGAGCAGC GAACCCGCAA AATTCGCCGA CAGCCTGCAA  
 1801 CCCGGCTTTA CCCTGCTCAA AACCGGCTAC GCCCTGACCG GCTACATCTC  
 1851 CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT  
 1901 TTACCGCACA GTTCCACCTC GCGGCCGAAC ACACCGCCCA CATCTTCCAA  
 1951 CACCTGCCCG AAACCGAACG CGACGACTTT CAGACAGCAC TGGATACACT  
 2001 GCGCGGCGAA CTCGACACCC TCCGCACCCA CAGCAGCGGA ACACAAAGCC  
 2051 ACATCCTCCT CCAACAGCTC CAACTCATCG CCCGACAGCT CGAACCCCTAC  
 2101 TACCGCGCCT ACCGCCAAAT TCCGCACAGG CAGCCCCAAA ATGCAGCCTG  
 2151 A

This corresponds to the amino acid sequence <SEQ ID 106; ORF19-1>:

1 MKTPLLKPLL ITS L P V F A S V F T A A S I V W Q L G E P K L A M P F V L G I I A G G L V D  
 51 L D N R L T G R L K N I I T T V A L F T L S S L T A Q S T L G T G L P F I L A M T L M T F G F T I L  
 101 G A V G L K Y R T F A F G A L A V A T Y T I L T Y T P E T Y W L T N P F M I L C G T V L Y S T A I L  
 45 151 L F Q I V L P H R P V Q E S V A N A Y D A L G G Y L E A K A D F F D P D E A A W I G N R H I D L A M  
 201 S N T G V I T A F N Q C R S A L F Y R L R G K H R H P R T A K M L R Y Y F A A Q D I H E R I S S A H  
 251 V D Y Q E M S E K F K N T D I I F R I H R L L E M Q G Q A C R N T A Q A L R A S K D Y V Y S K R L G  
 301 R A I E G C R Q S L R L L S D S N D S P D I R H L R R L L D N L G S V D Q Q F R Q L Q H N G L Q A E  
 351 N D R M G D T R I A A L E T S S L K N T W Q A I R P Q L N L E S G V F R H A V R L S L V V A A A C T  
 50 401 I V E A L N L N L G Y W I L L T A L F V C Q P N Y T A T K S R V R Q R I A G T V L G V I V G S L V P  
 451 Y F T P S V E T K L W I V I A S T T L F F M T R T Y K Y S F S T F F I T I Q A L T S L S L A G L D V  
 501 Y A A M P V R I I D T I I G A S L A W A V S Y L W P D W K Y L T L E R T A A L A V C S N G A Y L E  
 551 K I T E R L K S G E T G D D V E Y R A T R R R A H E H T A A L S S T L S D M S S E P A K F A D S L Q  
 601 P G F T L L K T G Y A L T G Y I S A L G A Y R S E M H E E C S P D F T A Q F H L A A E H T A H I F Q  
 55 651 H L P E T E P D D F Q T A L D T L R G E L D T L R T H S S G T Q S H I L L Q Q L Q L I A R Q L E P Y  
 701 Y R A Y R Q I P H R Q P Q N A A \*

Computer analysis of this amino acid sequence gave the following results:

Homology with predicted transmembrane protein YHFK of *H. influenzae* (accession number P44289)

ORF19 and YHFK proteins show 45% aa identity in 97 aa overlap:

60 orf19 6 LKPLLTSLPVFASVFTAASIVWQLGEPKLAMPFVLGGIIAGGLVDLDNXXTGRLKNIITT 65  
 L +I+++PVF +V AA +W +MP +LGIIAGGLVDLDN TGR LKN+ T  
 YHFK 5 LNAKVISTIPVFI AVNIAAVGIWFFDISSQSMPLILGGIIAGGLVDLDNRLTGRLKNVFFT 64



-110-

orf19 66 VALFTLSSLTAQSTLGTGLPFILAMTLMXXFTILGA 102  
 + F++SS Q +G + +I+ MT++T FT++GA  
 YHFK 65 LIAFSISSFIVQLHIGKPIQYIVLMTVLTFTFTMIGA 101

### 5 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF19 shows 92.2% identity over a 102aa overlap with an ORF (ORF19a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
10	orf19.pep	MKTPLLKPELLITSLPVFASVFTAASIVWQLGEPKLPFVLGIIAGGLVDLNDXXTGRLLK					
	orf19a	MKTPLLKPELLITSLPVFASVFTAASIVWQLGEPKLPFVLGIIAGGLVDLNDRLTGRLLK					
		10	20	30	40	50	60
15	orf19.pep	70	80	90	100		
		NIIITVALFTLSSLTAQSTLGTGLPFILAMTLMXXFTILGAX					
	orf19a	NIIATVALFTLSSLVAQSTLGTGLPFILAMTLMFTFGFTIMGAVGLKYRTFAFGALAVATY					
		70	80	90	100	110	120
20	orf19a	TTLTYTPETYWLTNPFMILCGTVLYSTAILLFIILPHRPVQENVANAYEALGSYLEAKA					
		130	140	150	160	170	180

The complete length ORF19a nucleotide sequence <SEQ ID 107> is:

	1	ATGAAAACCC	CACCCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
25	51	CGCCAGTGTC	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTG	GGCGAACCCA
	101	AGCTCGCCAT	GCCCTTCGTA	CTCGGCATCA	TCGCTGGCGG	CCTGGTCGAT
	151	TTGGACAACC	GCCTGACCGG	ACGGCTGAAA	AACATCATCG	CCACCGTCGC
	201	CCTGTTCAAC	CTCTCCTCAC	TTGTGCGGCA	AAGCACCTC	GGCAGAGTT
	251	TGCCATTGAT	CCTCGCCATG	ACCCTGATGA	CTTTCGGCTT	TACCATCATG
30	301	GGCGCGGTCG	GGCTGAAATA	CCGCACCTTC	GCCTTCGGCG	CACTCGCCGT
	351	CGCCACCTAC	ACCACACTTA	CCTACACCCC	CGAAACCTAC	TGGCTGACCA
	401	ACCCCTTTAT	GATTCTGTGC	GGAACCGTAC	TGTACAGCAC	CGCCATCATC
	451	CTGTTCCAAA	TCATCCTGCC	CCACCGCCCC	GTTCAAGAAA	ACGTGCGCAA
	501	CGCCTACGAA	GCACTCGGCA	GCTACCTCGA	AGCCAAAGCC	GACTTTTTTCG
35	551	ATCCCGACGA	AGCCGAATGG	ATAGGCAACC	GCCACATCGA	CCTCGCCATG
	601	AGCAACACCG	CGCTCATCAC	CGCCTTCAAC	CAATGCCGTT	CCGCCCTGTT
	651	TTACCGCCTT	CGCGGCAAAC	ACCGCCACCC	GCGCACCGCC	AAAATGCTGC
	701	GCTACTACTT	CGCCGCCCAA	GACATACACG	AACGCATCAG	CTCCGCCAC
	751	GTCGACTACC	AAGAGATGTC	CGAAAATTC	AAAAACACCG	ACATCATCTT
40	801	CCGCATCCAC	CGCTGCTCG	AAATGCAGGG	ACAAGCCTGC	CGCAACACCG
	851	CCCAAGCCCT	CGCGCAAGC	AAAGACTACG	TTTACAGCAA	ACGCCCTGGC
	901	CGCGCCATCG	AAGGCTGCCG	CCAATCGCTG	CGCCTCCTTT	CAGACAGCAA
	951	CGACAATCCC	GACATCCGCC	ACCTGCGCCG	CCTTCTCGAC	AACCTCGGCA
	1001	GCGTCGACCA	GCAGTTCGCG	CAACTCCAGC	ACAACGGCCT	GCAGGCAGAA
45	1051	AACGACCGCA	TGGCGGACAC	CCGCATCGCC	GCCCTCGAAA	CCGGCAGCCT
	1101	CAAAAACACC	TGGCAGGCAA	TCCGTCCGCA	GCTAAACCTC	GAATCCGGCG
	1151	TATTCCGCCA	TGCCGTCCGC	CTGTCCCTTG	TCGTTGCCGC	CGCCTGCACC
	1201	ATCGTCGAAG	CCCTCAACCT	CAACCTCGGC	TACTGGATAC	TACTGACCGC
	1251	CCTTTTCGTC	TGCCAACCCT	ACTACACCGC	CACCAAAGC	CGCGTCCGCC
50	1301	AGCGCATCGC	CGGCACCGTA	CTCGGCGTAA	TCGTGCGCTC	GCTCGTCCCC
	1351	TACTTTTACC	CCTCCGTCGA	AACCAAACTC	TGGATCGTCA	TCGCCAGTAC
	1401	CACCCCTCTT	TTCATGACCC	GCACCTACAA	ATACAGCTTC	TCGACATTTT
	1451	TCATCACCAT	TCAAGCCCTG	ACCAGCCTCT	CCCTCGCAGG	GTTGGACGTA
	1501	TACGCGGCCA	TGCCCGTACG	CATCATCGAC	ACCATTATCG	GCGCATCCCT
55	1551	TGCCTGGGCG	GCAGTCAGCT	ACCTGTGGCC	AGACTGGAAA	TACCTCACGC
	1601	TCGAACGCAC	CGCCGCCCTT	GCCGTATGCA	GCAACGGCGC	CTATCTCGAA
	1651	AAAATCACCG	AACGCCTCAA	AAGCGGCGAA	ACCGGCGACG	ACGTGCAATA
	1701	CCGCGCCACC	CGCCGCCGCG	CCCACGAACA	CACCGCCGCC	CTCAGCAGCA
	1751	CCCTTTCCGA	CATGAGCAGC	GAACCCGCAA	AATTCCGCCA	CAGCCTGCAA
60	1801	CCCGGCTTTA	CCCTGCTCAA	AACCGGCTAC	GCCCTGACCG	GCTACATCTC
	1851	CGCCCTCGGC	GCATACCGCA	GCGAAATGCA	CGAAGAATGC	AGCCCGGACT
	1901	TTACCGCACA	GTTCCACCTC	GCGCCGAAC	ACACCGCCCA	CATCTTCCAA
	1951	CACCTGCCCC	AAACCGAACC	CGACGACTTT	CAGACAGCAC	TGGATACACT
	2001	GCGCGGCGAA	CTCGACACCC	TCCGCACCCA	CAGCAGCGGA	ACACAAAGCC
65	2051	ACATCCTCCT	CCAACAGCTC	CAACTCATCG	CCCGGCAGCT	CGAACCCTAC
	2101	TACCGCGCCT	ACCGACAAAT	TCCGCACAGG	CAGCCCCAAA	ACGCAGCCTG
	2151	A				

This encodes a protein having amino acid sequence <SEQ ID 108>:

```

1  MKTPPLKPLL ITSPLVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
51  LDNRLTGRLL NIIATVALFT LSSLVAQSTL GTGLPFILAM TLMTEGFTIM
101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAI
5  151  LFQIILPHRP VQENVANAYE ALGSYLEAKA DFFDPDEAEW IGNRHIDLAM
201  SNTGVITAFN QCRSALFYRL RGKHRHPTA KMLRYFFAAQ DIHERISSAH
251  VDYQEMSEKF KNTDIIFRIH RLLEMQGOAC RNTAQALRAS KDYVYSKRLG
301  RAIEGCRQSL RLLSDSDNDP DIRHLRLLD NLGSVDQQFR QLQHNGLQAE
351  NDRMGDTRIA ALETGSLKNT WQAIRPQLNL ESGVFRHAVR LSLVVAAC
10  401  IVEALNLLNG YWILLTALFV CQPNYTATKS RVRQRIAGTV LGVIVGSLVP
451  YFTPSVETKL WIVIASTTLF FMTRTYKYSF STFFITIQL TSLSLAGLDV
501  YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLERTAAL AVCSNGAYLE
551  KITERLKSGE TGDDVEYRAT RRAHEHTAA LSSTLSDMSS EPAKFADSLQ
601  PGFTLLKGTG ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ
15  651  HLPETEPDDF QTALDTRLGE LDTLRTHSSG TQSHILLQQL QLIARQLEPY
701  YRAYRQIPHR QPQNAA*

```

ORF19a and ORF19-1 show 98.3% identity in 716 aa overlap:

```

20  orf19a.pep      10      20      30      40      50      60
    MKTPPLKPLLITSPLVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLLK
    orf19-1        10      20      30      40      50      60
    MKTPLLKPLLITSPLVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLLK

25  orf19a.pep      70      80      90      100     110     120
    NIIATVALFTLSSLVAQSTLGTGLPFILAMTLMTEGFTIMGAVGLKYRTFAFGALAVATY
    orf19-1        70      80      90      100     110     120
    NIITTVALFTLSSLTAQSTLGTGLPFILAMTLMTEGFTILGAVGLKYRTFAFGALAVATY

30  orf19a.pep      130     140     150     160     170     180
    TTLTYTPETYWLTNPFMILCGTVLYSTAIILFQIILPHRPVQENVANAYEALGSYLEAKA
    orf19-1        130     140     150     160     170     180
    TTLTYTPETYWLTNPFMILCGTVLYSTAILLFQIVLPHRPVQESVANAYDALGGYLEAKA

35  orf19a.pep      190     200     210     220     230     240
    DFFDPDEAEWIGNRHIDLAMSNTGVITAFNQCRSALFYRLRGKHRHPTAKMLRYFFAAQ
    orf19-1        190     200     210     220     230     240
    DFFDPDEAAWIGNRHIDLAMSNTGVITAFNQCRSALFYRLRGKHRHPTAKMLRYFFAAQ

40  orf19a.pep      250     260     270     280     290     300
    DIHERISSAHVDYQEMSEKFNTDIIFRIHRLLEMQGOACRNTAQALRASKDYVYSKRLG
    orf19-1        250     260     270     280     290     300
    DIHERISSAHVDYQEMSEKFNTDIIFRIHRLLEMQGOACRNTAQALRASKDYVYSKRLG

45  orf19a.pep      310     320     330     340     350     360
    RAIEGCRQSLRLLSDSDNDPDIRHLRLLDNLGSVDQQFRQLQHNGLQAENDRMGDTRIA
    orf19-1        310     320     330     340     350     360
    RAIEGCRQSLRLLSDSDNDSPDIRHLRLLDNLGSVDQQFRQLQHNGLQAENDRMGDTRIA

50  orf19a.pep      370     380     390     400     410     420
    ALETGSLKNTWQAIRPQLNLGSGVFRHAVRSLVVAACACTIVEALNLLNGYWILLTALFV
    orf19-1        370     380     390     400     410     420
    ALETSSLKNTWQAIRPQLNLGSGVFRHAVRSLVVAACACTIVEALNLLNGYWILLTALFV

55  orf19a.pep      430     440     450     460     470     480
    CQPNYTATKSRVRQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFEMTRTYKYSF
    orf19-1        430     440     450     460     470     480
    CQPNYTATKSRVRQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFEMTRTYKYSF

60  orf19a.pep      490     500     510     520     530     540
    STFFITIQLTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLETAAL
    orf19-1        490     500     510     520     530     540
    STFFITIQLTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLETAAL

```

-112-

	orf19-1	STFFITIQALTSLSLAGLDVYAAMPVRIIDTTIIGASLAWAAVSYLWPDWKYLTLERTAAL	490	500	510	520	530	540
5			550	560	570	580	590	600
	orf19a.pep	AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSDMSSEPAKFADSLQ						
	orf19-1							
		AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSDMSSEPAKFADSLQ	550	560	570	580	590	600
10			610	620	630	640	650	660
	orf19a.pep	PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLLAAEHTAHIFQHLPETEPDDF						
	orf19-1							
		PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLLAAEHTAHIFQHLPETEPDDF	610	620	630	640	650	660
15			670	680	690	700	710	
	orf19a.pep	QTALDTLRGELDTLRTHSSGTQSHILLQQLLIARQLEPYRAYRQIPHROPQNAAX						
	orf19-1							
		QTALDTLRGELDTLRTHSSGTQSHILLQQLLIARQLEPYRAYRQIPHROPQNAAX	670	680	690	700	710	
20			670	680	690	700	710	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF19 shows 95.1% identity over a 102aa overlap with a predicted ORF (ORF19.ng) from *N.*

*gonorrhoeae*:

25	orf19.pep	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLK	60
	orf19ng	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK	60
30	orf19.pep	NIIITVALEFTLSSSLTAQSTLGTGLPFILAMTMTXXFTILGAX	103
	orf19ng	NIIATVALEFTLSSSLTAQSTLGTGLPFILAMTMTFGFTILGAVGLKYRTFAFGALAVATY	120

An ORF19ng nucleotide sequence <SEQ ID 109> is predicted to encode a protein having amino acid sequence <SEQ ID 110>:

35	1	MKTPLLKPLL	ITSLPVFASV	FTAASIVWQL	GEPKLAMPFV	LGIAGGLVD
	51	LDNRLTGRLE	NIIATVALEF	LSSSLTAQSTL	GTGLPFILAM	TMTFGFTIL
	101	GAVGLKYRTF	AFGALAVATY	TTLTYTPETY	WLTNPFMILC	GTVLYSTAI
	151	LFQIILPHRP	VQESVANAYE	ALGGYLEAKA	DEFDPEAAW	IGNRHIDLAM
	201	SNTGVITAFN	QCRSALFYRL	RGKRRHERTA	KMLRYFFAAQ	DIHERISSAH
	251	VDYQEMSEKF	KNTDIIIFRIR	RLLEMQGQAC	RNTAQAIRSG	KDYVYSKRLG
40	301	RAIEGCRQSL	RLSDGNDSP	DIRHLSRLLD	NLGSVDQQFR	QLRHSDSPAE
	351	NDRMGDTRIA	ALETGSFKNT	*		

Further work revealed the complete nucleotide sequence <SEQ ID 111>:

45	1	ATGAAAACCC	CACTCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
	51	CGCCAGTGTC	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTA	GGCGAACCCA
	101	AGCTCGCCAT	GCCCTTCGTA	CTCGGCATCA	TCGCCGGCGG	CCTGGTCGAT
	151	TTGGACAACC	GCCTGACCGG	ACGGCTGAAA	AACATCATCG	CCACCGTCGC
	201	CCTGTTTACC	CTCTCCTCGC	TCACGGCGCA	AAGCACCTTC	GGCACAGGGC
	251	TGCCCTTCAT	CCTCGCCATG	ACCCTGATGA	CCTTCGGCTT	TACCATTTTA
50	301	GGCGCGGTCG	GGCTGAAATA	CCGCACCTTC	GCCTTCGGCG	CACTCGCCGT
	351	CGCCACCTAC	ACCACGCTTA	CCTACACCCC	CGAAACCTAC	TGGCTGACCA
	401	ACCCCTTCAT	GATTTTATGC	GGCACCGTAC	TGTACAGCAC	CGCCATCATC
	451	CTGTTCCAAA	TCATCCTGCC	CCACCGCCCC	GTCCAAGAAA	GGCTCGCCAA
	501	TGCCTACGAA	GCACTCGGCG	GCTACCTCGA	AGCCAAAGCC	GACTTCTTCG
	551	ACCCGATGA	GGCAGCTGG	ATAGGCAACC	GCCACATCGA	CCTCGCCATG
55	601	AGCAACACCG	GCGTCATCAC	CGCCTTCAAC	CAATGCCGTT	CCGCCCTGTT
	651	TTACCGTTTG	CGCGGCAAAC	ACCGCCACCC	GCGCACCGCC	AAAATGCTGC
	701	GCTACTACTT	CGCCGCCCAA	GACATCCACG	AACGCATCAG	CTCCGCCAC
	751	GTCGACTACC	AAGAGATGTC	CGAAAAATTC	AAAAACACCG	ACATCATCTT
60	801	CCGCATCCGC	CGCCTGCTCG	AAATGCAGGG	GCAGGCGTGC	CGCAACACCG
	851	CCCAAGCCAT	CCGGTCGGGC	AAAGACTAcg	tTTACAGCAA	ACGCCTCGGA
	901	CGCGCCATcg	aaggctgCCG	CCAGTCGctg	cgcctCCTTt	cagacggcaA
	951	CGACAGTCCC	GACATCCGCC	ACCTGAGccg	CCTTCTCGAC	AACCTCGgca

-113-

1001 GCGTcgacca gcagtTCgc caactCCGAC ACAGcgactC CCCCCcgaa  
 1051 Aacgaccgca tgggcgacaC CCGCATCGCC GCCCtcgaaa ccggcagctT  
 1101 caaaaaCAcc tggcaggCAA TCCGTCCGCa gctgaaCCTC GAATCatgCG  
 1151 TATTCCGCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC  
 1201 ATCGTCgaag cCCTCAACCT CAACCTCGGC TACTGGATAC TGCTGACCGC  
 1251 CTTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAAGC CGCGTGTACC  
 1301 AACGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC  
 1351 TACTTCACCC CCTCCGTGCA AACCAAATC TGGATGTCA TCGCCGTAC  
 1401 CACCCTGTTT TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT  
 1451 TCATCACCAT TCAGGCACGT ACCAGCCTCT CCCTCGCAGG TTTGGACGTA  
 1501 TACGCCGCCA TGCCCGTGCG CATCATcgaC ACCATTATCG GCGCATCCCT  
 1551 TGCCTGGGCG GCGGTGAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC  
 1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAGCGGCAC ATACCTCCAA  
 1651 AAAATTGCCG AACGCCTCAA AACCGGCGAA ACCGGCGACG ACATAGAATA  
 1701 CGCATCACC CGCCGCCGCG CCCACGAACA CACCGCGGCC CTCAGCAGCA  
 1751 CCCTTTCCGA CATGAGCAGC GAACCGCAA AATTCGCCGA CAGCCTGCAA  
 1801 CCCGGCTTTA CCCTGCTCAA AACCGGCTAC GCCCTGACCG GCTACATCTC  
 1851 CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGATGC AGCCCGGACT  
 1901 TTACCGCACA GTTCCACCTT GCCGCCGAAC ACACCGCCCA CATCTTCAA  
 1951 CACCTGCCCG ACATGGGACC CGACGACTTT CAGACGGCAT TGGATACACT  
 2001 GCGCGGCGAA CTCGGCACCC TCCGCACCCG CAGCAGCGGA ACACAAAGCC  
 2051 ACATCTCCT CCAACAGCTC CAACTCATCG CccgGCAACT CGAACCTAC  
 2101 TACCGCGCCT ACCGACAAAT TCCGCACAGG CAGCCCCAAA ACGCAGCCTG  
 2151 A

25 This corresponds to the amino acid sequence <SEQ ID 112; ORF19ng-1>:

1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD  
 51 LDNRLTGRLE NIIATVALFT LSSSLTAQSTL GTGLPFILAM TLMFTGFTIL  
 101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAIL  
 151 LFQIILPHRP VQESVANAYE ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM  
 201 SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYFAAQ DIHERISSAH  
 251 VDYQEMSEKF KNTDIIIFRIR RLLEMQGQAC RNTAQAIRSG KDYVYSKRLG  
 301 RAIEGCRQSL RLLSDGNDSP DIRHLSRLLD NLGSVDQQFR QLRHSDSPA  
 351 NDRMGDTRIA ALETGSFKNT WQAIRPQLNL ESCVFRHAVR LSLVVAAC  
 401 IVEALNINLG YWILLTALFV CQPNYTATKS RVYQRIAGTV LGVIVGSLVP  
 451 YFTPSVETKL WIVIAGTTLF FMTRTYKYSF STFFITIQAL TSLSLAGLDV  
 501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLETAAL AVCSSGTYLQ  
 551 KIAERLKTGE TGDDIEYRIT RRRRAHEHTAA LSSTLSDMSS EPAKFADSLQ  
 601 PGFTLLKTGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ  
 651 HLPDMGPDDF QTALDTRLGE LGTLRTRSSG TQSHILLQQL QLIARQLEPY  
 701 YRAYRQIPHR QPQNAA\*

ORF19ng-1 and ORF19-1 show 95.5% identity in 716 aa overlap:

		10	20	30	40	50	60
45	orf19-1.pep	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLE					
	orf19ng-1	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLE					
		10	20	30	40	50	60
50	orf19-1.pep	70	80	90	100	110	120
	orf19ng-1	NIIATVALFTLSSSLTAQSTLGTGLPFILAMTLMFTGFTILGAVGLKYRTFAFGALAVATY					
		70	80	90	100	110	120
55	orf19-1.pep	130	140	150	160	170	180
	orf19ng-1	TTLTYTPETYWLTNPFMILCGTVLYSTAILLFQIILPHRPVQESVANAYDALGGYLEAKA					
		130	140	150	160	170	180
60	orf19-1.pep	190	200	210	220	230	240
	orf19ng-1	DFFDPDEAAWIGNRHIDLAMSNITGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYFAAQ					
		190	200	210	220	230	240
65	orf19-1.pep	250	260	270	280	290	300
		DIHERISSAHVDYQEMSEKFKNTDIIIFRIHRLLEMQGQACRNTAQALRASKDYVYSKRLG					

In addition, ORF19ng-1 shows significant homology to a hypothetical gonococcal protein previously entered in the databases:

sp|O333369|YOR2\_NEIGO HYPOTHETICAL 45.5 KD PROTEIN (ORF2) gnl|PID|e1154438  
(AJ002423) hypothetical protein [Neisseria gonorrh] Length = 417  
Score = 1512 (705.6 bits), Expect = 5.3e-203, P = 5.3e-203  
Identities = 301/326 (92%), Positives = 306/326 (93%)

Query: 307 RQSLRLSDGNDSPDIRHLSRLLDNLGSDVQQFRQLRHSDSPAENDRMGDTRIAALETGS 366

```
Query:      507  RQSLRLLSDGNDSDIRHLRLLLDNLGSVDQQFQRLRHSDSPAENDRMGDTRIAALETGS 509
           508  RQSLRLLSDGNDSDIRHLRLLLDNLGSVDQQFQRLRHSDSPAENDRMGDTRIAALETGS 510
Sbjct:      1    RQSLRLLSDGNDSDIRHLRLLLDNLGSVDQQFQRLRHSDSPAENDRMGDTRIAALETGS 60
```

Query: 367 FKNTWQAIRPQLNLESCVFRHAVRLSLVVAAACTIVEALNINLGYWILLTALFVCQPNYT 426

Subjct: 61 FKNTWQAIRPOLNLESGVFRHAVRLSLVVAACTIVEALNLNGLGYWILLTRLFVCQPNYT 120

Query: 427 ATKSRVYORIAGTVLGVIVGSLVPYFTPSVETKLWIVIAGTTLFFMTRTYKYSESTFFIT 486

Query: 427 ATKSRVYQRIAGTVLGVIVGSLVPYFTSPVETKLWIVIAGTTLFFMTRTYKYSFSTFFIT 180  
 Spict: 121 ATKSRVYQRIAGTVLGVIVGSLVPYFTSPVETKLWIVIAGTTLFFMTRTYKYSFSTFFIT 180

Query: 487 IOALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLERTAALAVCSSG 546

Query: 487 IQALTSLSLAGLDVYAAMPVRIIDTTIIGASLAWAAVSYLWPDWKYLT LERTAALAVCSSG 516  
 Spict: 181 IQALTSLSLAGLDVYAAMPVRIIDTTIIGASLAWAAVSYLWPDWKYLT LERTAALAVCSSG 240

Query: 547 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSSTLSDMSSEPAKFADSLQPGFTLL 606  
 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSSTLSDMSSEPAKFAD+ P  
 Sbjct: 241 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSSTLSDMSSEPAKFADTCNPALPCS 300

5 Query: 607 KTGyALTGYISALGAYRSEMHEECSP 632  
 K ALTGYISALG ++ + +P  
 Sbjct: 301 KPATALTYISALGHTAAKCTKNAAP 326

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein (the first of which is also seen in the meningococcal protein), and on homology with the YHFK protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 14

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 113>:

```

15      1  ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTG
      51  GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCGGCG
     101  CGGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAAAC T GCCCAACCTG
     151  CTTCGCCGCG TGTTGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
     201  TTTGGCGGAA TACAAGGAAA CGCGTTCAAA AGAGGCGG.C GAAGCCTTA
     251  TCCGCCATGT GCGGGGATG CTGTCGTTT TACTGGTTAT CGTTACCGCG
     301  CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGAGTT
     351  TTGCCCAAGA TGCCGACAAA TTTCAGCTCT CCATCGATT GCTGCGGATT
     401  ACGTTTCCTT ATATATTATT GATTTCCTG TCTTCATTG TCGGCTCGGT
     451  ACTCAATTCT TATCATAAGT TCGGCATTCC GCGGTTTACG CCAC.GTTTC
     501  TGAACGTGTC GTTTATCGTA TTCGCGCTGT TTTTCGTGCC GTATTTTCGAT
     551  CCGCCCGTTA CCGCGCGyGGC GTGGGCGGTC TTTGTGCGCG GCATTTTGCA
     601  ACTCGmTTC CAACTGCCCT GGCTGGCGAA ACTGGGCTTT TTGAAACTGC
     651  CCAAACtGAG TTTCAAAGAT GCGGCGGTCA ACCGCGTGAT GAAACAGATG
     701  GCGCCTGCgA TTTTgGGCGT GAgCGTGGCG CAGGTTTCTT TGGTGATCAA
     751  CACGATTTTc GCGTCTTATC TGCAATCGGG CAGCGTTTCA TGGATGTATT
     801  ACGCCGACCG CATGATGGAG CTGCCAGCG GCGTGCTGGG GGCGGCACTC
     851  GGTACGATTT TGCTGCCGAC TTGTGCCAAA CACTCGGCAA ACCaAGATAC
     901  GGaACAGTTT TCCGCCCTGC TCGACTGGGG TTTGCGCCTG TGCATGCTgc
     951  TGACGCTGCC GGCGgcGGTc GGA CTGGCGG TGTGTCTGTT cCCgCtGGTg
    1001  GCGACGCTGT TTATGTACCG CGWATTTACG CTGTTTGACG CGCAGATGAC
    1051  GCAACACGCG CTGATTGCCT ATTCTTTCGG TTTAATCGGC TTAATCATGA
    1101  TTAAAGTGTT GGCACCCGGC TTCTATGCGC GGCAAAACAT CAAwAmGCCC
    1151  GTCAAATCG CCATCTTCAC GCTCATCTGC mCGCAGTTGA TGAACCTTGs
    1201  CTTTAYCGGC CACTrrAAC rCasTCGGAC TTTGCTTGC CATCGGCTG
    1251  GCGCGGTGTA TCAATGCCGG ATTGTTGTTT TACCTGTTGC GCAGACACGG
    1301  TATTTACCAA CCTGG.CAAG GGTGGGCAG CGTTCTT.AG CAAAAATGCT
    1351  GcTCTCGCTC GCCGTGA
  
```

This corresponds to the amino acid sequence <SEQ ID 114; ORF20>:

```

45      1  MNMLGALAKV GSLTMVSRVL GFVRDTVIAF AFGAGMATDA FFVAFKLPNL
      51  LRRVFAEGAF AQAFVPILAE YKETRSKEAX EAFIRHVAGM LSFVLVIVTA
     101  LGILAAPWVI YVSAPSFAQD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
     151  LNSYHKFGIP AFTPXFLNVS FIVFALFFVP YFDPPTVAXA WAVFVGGILQ
     201  LXFQLPWLA LGLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN
     251  TIFASYLQSG SVSWMYADR MMELPSGVLG AALGTILLPT LSKHSANQDT
     301  EQFSALLDWG LRLCMLLTLP AAVGLAVLSF PLVATLFMYR XFTLFDAQMT
     351  QHALIAYSFG LIGLIMIKVL APGFYARQNI XXPVKIAIFT LICXQLMNLX
     401  FXGPLXXIGL SLAIGLGACI NAGLLFYLLR RHGIYQPXQG LGSVLXQKCC
     451  SRSP*
  
```

These sequences were elaborated, and the complete DNA sequence <SEQ ID 115> is:

```

55      1  ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTG
      51  GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCGGCG
  
```

101 CGGGTATGGC GACGGATGCG TTTTTTGTCT CGTTCAAACCT GCCCAACCTG  
 151 CTTCCGCCGCG TGTTCGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT  
 201 TTTGGCGGAA TACAAGGAAA CGCGTTCAAA AGAGCGCGCG GAGGCTTTTA  
 251 TCCGCCATGT GCGGGGATG CTGTCGTTG TACTGGTTAT CGTTACCGCG  
 301 CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGGTTT  
 351 TGCCCAAGAT GCCGACAAAT TTCAGCTCTC CATCGATTG CTGCGGATTA  
 401 CGTTTCCTTA TATATTATG ATTTCCCTGT CTTCAATTGT CGGCTCGGTA  
 451 CTCAATTCTT ATCATAAGTT CGGCATTCCG GCGTTTACGC CCACGTTTCT  
 501 GAACGTGTCG TTTATCGTAT TCGCGCTGTT TTTCTGCGG TATTTCGATC  
 551 CGCCCGTTAC CGCGCTGGCG TGGGCGGTCT TTGTGCGCGG CATTTTGCAA  
 601 CTCGGCTTCC AACTGCCCTG GCTGGCGAAA CTGGGCTTTT TGAACCTGCC  
 651 CAAACTGAGT TTCAAAGATG CGGCGTCAA CCGCGTGATG AAACAGATGG  
 701 CGCCTGCGAT TTTGGGCGTG AGCGTGGCGC AGGTTTCTTT GGTGATCAAC  
 751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTATTA  
 801 CGCCGACCGC ATGATGGAGC TGCCACGCGG CGTGCTGGGG GCGCACTCG  
 851 GTACGATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAA CCAAGATACG  
 901 GAACAGTTT CCGCCCTGCT CGACTGGGGT TTGCGCTGT GCATGCTGCT  
 951 GACGCTGCGG GCGGCGGTCT GACTGGCGGT GTTGTCTGTT CCGCTGGTGG  
 1001 CGACGCTGTT TATGTACCGC GAATTACGC TGTTCGACG CGCATGACG  
 1051 CAACACGCGC TGATTGCTA TTCTTTCGGT TTAATCGGCT TAATCATGAT  
 1101 TAAAGTGTG GCACCCGGCT TCTATGCGCG GCAAAACATC AAAACGCCCG  
 1151 TCAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTTGCC  
 1201 TTTATCGGCC CACTGAAACA CGTCGGAAT TCGCTTGCCA TCGGTCTGGG  
 1251 CGCGTGTATC AATGCCGAT TGTGTTTGA CCTGTGCGC AGACACGTA  
 1301 TTTACCAACC TGGCAAGGT TGGGACGCGT TCTTAGCAA AATGCTGCTC  
 1351 TCGCTCGCCG TGATGTGCGG CGGACTGTGG GCAGCGCAGG CTTACCTGCC  
 1401 GTTTGAATGG GCGCACGCGG GCGGAATGCG GAAAGCGGG CAGCTCTGCA  
 1451 TCCTGATTGC CGTCGCGGCG GGAAGTATT TCGCATCACT GCGGCTTTG  
 1501 GGCTTCCGTC CGGCCATT TCAAACGCGT GAAACTGA

30 This corresponds to the amino acid sequence <SEQ ID 116; ORF20-1>:

1 MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL  
 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAA EAFIRHVAGM LSFVLVIVTA  
 101 LGILAAPWVI YVSAPGFAQD ADKFQLSIDL LRITFPYILL ISLSFVGSV  
 151 LNSYHKFGIP AFTPTFLNVS FIVFALFFVP YFDPVPTALA WAVFVGILQ  
 201 LGFQLPWLAK LGFLKLPKLS FKDAVNRVM KQMAPAILGV SVAQVSLVIN  
 251 TIFASYLQSG SVSWMYADR MMELPSGVLG AALGTILLPT LSKHSANQDT  
 301 EQFSALLDWG LRLCMLTLP AAVGLAVLSF PLVATLFMYR EFTLFDAQMT  
 351 QHALIAYSFG LIGLIMIKVL APGFYARQNI KTPVKIAIFT LICTQLMNL  
 401 FIGPLKHVGL SLAIGLGACI NAGLLEFYLLR RHGIYQPGKG WAAFLAKMLL  
 451 SLAVMCGGLW AAQAYLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAAL  
 501 GFRPRHFKRV EN\*

Computer analysis of this amino acid sequence gave the following results:

Homology with the MviN virulence factor of *S. typhimurium* (accession number P37169)

ORF20 and MviN proteins show 63% aa identity in 440aa overlap:

45 Orf20 1 MNMLGALAKV GSLTMVSRVL GFVRDVIARA FGAAGMATDA FFVAFKLPNL LRRVFAEGAF 60  
 MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDA FFVAFKLPNL LRR+FAEGAF  
 MviN 14 MNLLKSLAAVSSMTMF SRVLGFARDAIVARIFGAGMATDA FFVAFKLPNL LRRIFAEGAF 73  
 50 Orf20 61 AQAFVPILAEYKETRSKEAXEAFIRHVAGM LSFVLVIVTALGILAAPWVIYVSAPGFAQD 120  
 +QAQFVPILAEYK + +EA F+ +V+G+L+ L +VT G+LAAPWVI V+AP FA  
 MviN 74 SQAQFVPILAEYKSKQGEAETRIFVAYVSGLLTLALAVVTVAGMLAAPWVIMVTAPGFADT 133  
 55 Orf20 121 ADKFQLSIDLLRITFPYILLISLSFVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP 180  
 ADKF L+ LLRITFPYILLISL+S VG++LN+++F IPAF P FLN+S I FALF P  
 MviN 134 ADKFALTQLLRITFPYILLISLASLVGAILNTWNRESIPAFAPTFLNISMIGFALFAAP 193  
 60 Orf20 181 YFDPVPTAXAWAVFVGILQLXFLPWLAKLGFLKLPKLSFKDAVNRVMKQMAPAILGV 240  
 YF+PPV A AWAV VGG+LQL +QLP+L K+G L LP+++F+D RV+KQMAPAILGV  
 MviN 194 YFNPPVLALAWAVTVGGVLQVLVYQLPYLKKIGMLVLPINFRDGTAMRVVKQMGPAAILGV 253  
 Orf20 241 SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT 300  
 SV+Q+SL+INTIFAS+L SGVSWMYADRME PSGVLG ALGTILLP+LSK A+ +  
 MviN 254 SVSISLIINTIFASFLASGVSWMYADRLMEFSPGVLGVALGTILLPSLSKSFASGNH 313

-117-

Orf20 301 EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHAIAYSFG 360  
 +++ L+DWGLRLC LL LP+AV L +L+ PL +LF Y FT FDA MTQ ALIAYS G  
 MviN 314 DEYCRIMDWGLRLCFLALPSAVALGILAKPLTVSLFQYGKFTAFDAAMTQRALIAYSVG 373

5 Orf20 361 LIGLIMIKVLAPGFYARQNIXXPVKIAIFTLICXQLMNLXFXXXXXXXXXXXXXXXXXXXCI 420  
 LIGLI++KVLAPGFY+RQ+I PVKIAI TLI QLMNL F C+  
 MviN 374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433

10 Orf20 421 NAGLLFYLLRRHGIYQXPXQG 440  
 NA LL++ LR+ I+ P G  
 MviN 434 NASLLYWQLRKQNIPTPQPG 453

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF20 shows 93.5% identity over a 447aa overlap with an ORF (ORF20a) from strain A of *N.*

### meningitidis:

	10	20	30	40	50	60
orf20.pep	MNMLGALAKVGSLTMVSRVLG	FVRD	TVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF			
orf20a	MNMLGALVKVGSLTMVSRVLG	FVRD	TVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF			
	10	20	30	40	50	60
orf20.pep	70	80	90	100	110	120
orf20a	70	80	90	100	110	120
	130	140	150	160	170	180
orf20.pep	ADKFQLSIDLRLITFPYILLISLSSEFVGSVLNSYHKFGIPAF	TPXFLNVSFIVFALFFVP				
orf20a	ADKFQLSIDLRLITFPYILLISLSSEFVGSVLNSYHKFSIPAF	TPXFLNVSFIVFALFFVP				
	130	140	150	160	170	180
orf20.pep	190	200	210	220	230	240
orf20a	190	200	210	220	230	240
	250	260	270	280	290	300
orf20.pep	SVAQVSLVINTIFASYLQSGSVSWMYYADRMELPSGVLGAALGTILLPTLSKHSANQDT					
orf20a	SVAQVSLVINTIFASYLQSGSVSWMYYADRMELPGGVLGAALGTILLPTLSKHSANQDT					
	250	260	270	280	290	300
orf20.pep	310	320	330	340	350	360
orf20a	310	320	330	340	350	360
	370	380	390	400	410	420
orf20.pep	LIGLIMIKVLAPGFYARQNIXXPVKIAIFTLICXQLMNLXFXGPLXXIGLSLAIGLGACI					
orf20a	LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICXQLMNLAFIGPLKHVGLSLAIGLGACI					
	370	380	390	400	410	420
orf20.pep	430	440	450			
orf20a	430	440	450			
	430	440	450	460	470	480

The complete length ORF20a nucleotide sequence <SEQ ID 117> is:

1 ATGAATATGC TGGGAGCTTT GGTAAAAGTC GGCAGCCTGA CGATGGTGTG  
 51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGC GCATTGCGCG  
 101 CAGGCATGGC GACGGATGCG TTCTTGTGCG CGTTCAACT GCCCAACCTG



-118-

151 CTTCCGCGCG TGTTCGCGGA GGGGCGGTTT GCCAAGCGT TTGTGCCGAT  
 201 TTTGGCGGAA TATAAGGAAA CGCGTTCTAA AGAGGCGACG GAGGCTTTTA  
 251 TCCGCCATGT GGCGGGGATG CTGTCGTTTG TACTGGTCAT CGTTACCGCG  
 301 CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGGTTT  
 351 TGCCAAAGAT GCCGACAAAT TTCAGCTCTC TATCGATTG CTGCGGATTA  
 401 CGTTTCCTTA TATCTTATTG ATTTCACTTT CCTCTTTTGT CGGCTCGGTA  
 451 CTCAATTCTT ATCATAAATT CAGCATTCTT GCGTTTACGC CCACGTTTCT  
 501 GAACGTGTCT TTTATCGTAT TCGCGCTGTT TTTCGTGCCG TATTTTCGATC  
 551 CTCCCGTTAC CGCGCTGGCT TGGGCGGTTT TTGTCGGCGG CATTTTGCAA  
 601 CTCGGCTTCC AACTGCCCTG GCTGGCGAAA CTGGGTTTTT TGAAACTGCC  
 651 CAAACTGAGT TTCAAAGATG CGGCGGTCAA CCGCGTGATG AAACAGATGG  
 701 CGCCTGCGAT TTTGGGCGTG AGCGTGGCGC AGATTTCTTT GGTGATCAAC  
 751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTATTA  
 801 CGCCGACCGC ATGATGGAAC TGCCCGGCGG CGTGCTGGGG GCGGCACTCG  
 851 GTACGATTTT GCTGCCGACT TTGTCCAAC ACTCGGCAA CCAAGATACG  
 901 GAACAGTTTT CCGCCCTGCT CGACTGGGGT TTGCGCNTGT GCATGCTGCT  
 951 GACGCTGCCG GCGGCGGTCT GAATGGCGGT GTTGTCTGTC CCGCTGGTGG  
 1001 CAACCTTGTT TATGTACCGA GAATTCACGC TGTTTGACGC GCAGATGACG  
 1051 CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGTT TAATCATGAT  
 1101 TAAAGTGTG GCGCCCGGCT TTTATGCGCG GCAAAACATC AAAACGCCCG  
 1151 TCAAAATCGC CATCTTCACG CTCATTGTCG CGCAGTTGAT GAACCTTGCC  
 1201 TTTATCGGCC CACTGAAACA CGTCGGACTT TCGCTTGCCA TCGGTCTGGG  
 1251 CGCGTGTATC AATGCCGGAT TGTTGTTTTA CCTGTTGCGC AGACACGGTA  
 1301 TTTACCAACC TGGAAGGGT TGGGCGCGT TCTTGGCAAA AATGCTGCTC  
 1351 TCGTCGCGC TGATGGGAGG CGGCCTGTAT GCCGCCCAA TCTGGCTGCC  
 1401 GTTCGACTGG GCACACGCCG GCGGAATGCA AAAGGCCGCC CGGCTCTTCA  
 1451 TCCTGATTGC CGTCGGCGGC GGACTGTATT TCGCATCACT GCGGCTTTG  
 1501 GGCTCCGTC CGGCCCATTT CAAACGCGT GAAAGCTGA

This encodes a protein having amino acid sequence <SEQ ID 118>:

30 1 MNMLGALVKV GSILTMVSRVL GFVRDVIAR AFGAGMATDA FFAVFKLPNL  
 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLVIUTA  
 101 LGILAAPWVI YVSAPGFAKD ADKFQLSIDL LRITFPYILL ISLSSFVGSV  
 151 LNSYHKFSIP AFTPTFLNVS FIVFALFFVP YFDPPTALTA WAVFVGILQ  
 201 LGFLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQISLVIN  
 35 251 TIFASYLQSG SVSWMYADR MMELPGGVLG AALGTILLPT LSKHSANQDT  
 301 EQFSALLDWG LRCMLLTLP AAVGMAVLSF PLVATLFMYR EFTLFDAQMT  
 351 QHALIAYSFG LIGLIMIKVL APGFYARQNI KTPVKIAIFT LICTQLMNL  
 401 FIGPLKHVGL SLAIGLGACI NAGLLFYLLR RHGIYQPGKG WAAFLAKMLL  
 451 SLAVMGGGLY AAQIWLFPDW AHAGGMQKAA RLFILIAVGG GLYFASLAAL  
 40 501 GFRPRHFKRV ES\*

ORF20a and ORF20-1 show 96.5% identity in 512 aa overlap:

		10	20	30	40	50	60
orf20a.pep		MNMLGALVKV	GSILTMVSRVL	GFVRDVIAR	AFGAGMATDA	FFVAFKLPNL	LRRVFAEGAF
45 orf20-1		MNMLGALVKV	GSILTMVSRVL	GFVRDVIAR	AFGAGMATDA	FFVAFKLPNL	LRRVFAEGAF
		70	80	90	100	110	120
orf20a.pep		AQAFVPILAE	YKETRSKEAT	EAFIRHVAGM	LSFVLVIUTA	LGILAAPWVI	YVSAPGFAKD
50 orf20-1		AQAFVPILAE	YKETRSKEAT	EAFIRHVAGM	LSFVLVIUTA	LGILAAPWVI	YVSAPGFAKD
		130	140	150	160	170	180
orf20a.pep		ADKFQLSIDL	LRITFPYILL	ISLSSFVGSV	LNSYHKFSIP	AFTPTFLNVS	FIVFALFFVP
55 orf20-1		ADKFQLSIDL	LRITFPYILL	ISLSSFVGSV	LNSYHKFSIP	AFTPTFLNVS	FIVFALFFVP
		190	200	210	220	230	240
orf20a.pep		YFDPPTALTA	WAVFVGILQ	LGFLPWLAK	LGFLKLPKLS	FKDAAVNRVM	KQMAPAILGV
60 orf20-1		YFDPPTALTA	WAVFVGILQ	LGFLPWLAK	LGFLKLPKLS	FKDAAVNRVM	KQMAPAILGV
		250	260	270	280	290	300
orf20a.pep		SVAQISLVIN	TIFASYLQSG	SVSWMYADR	MMELPGGVLG	AALGTILLPT	LSKHSANQDT

-119-

	orf20-1	SVAQVSLVINTIFASYLQSGSVSWMYADRMMLPSGVLGAALGTILLPTLSKHSANQDT 250 260 270 280 290 300
5	orf20a.pep	EQFSALLDWGLRXCMLLTLPAAVGMAVLSPFLVATLFMYREFTLFDAQMTQHALIAYSFG 310 320 330 340 350 360
	orf20-1	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG 310 320 330 340 350 360
10	orf20a.pep	LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI 370 380 390 400 410 420
	orf20-1	LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI 370 380 390 400 410 420
15	orf20a.pep	NAGLLFYLLRRHGIYPQPGKWAAFLAKMLLSLAVMGGGLYAAQIWLPFDWAHAGGMQKAA 430 440 450 460 470 480
	orf20-1	NAGLLFYLLRRHGIYPQPGKWAAFLAKMLLSLAVMCGGLWAAQAYLPFEWAHAGGMRKAG 430 440 450 460 470 480
20	orf20a.pep	RLFILIAVGGGLYFASLAALGFRPRHFKRVESX 490 500 510
	orf20-1	QLCILIAVGGGLYFASLAALGFRPRHFKRVENX 490 500 510

### Homology with a predicted ORF from *N.gonorrhoeae*

30 ORF20 shows 92.1% identity over a 454aa overlap with a predicted ORF (ORF20ng) from *N. gonorrhoeae*:

	orf20.pep	MNMLGALAKVGS LTMVSRVLGFVRD TVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF 	60
35	orf20ng	MNMLGALAKVGS LTMVSRVLGFVRD TVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF 	60
	orf20.pep	AQAFVPILA EYKETRSKEAXEAFIRHVAGMLS FVLVIVTALGILAAPWVIYVSAPSFAQD 	120
	orf20ng	AQAFVPILA EYKETRSKEATEAFIRHVAGMLS FVLVIVTALGILAAPWVIYVSAPGFTKD 	120
40	orf20.pep	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP 	180
	orf20ng	ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFPTPTFLNISFIVFALFFVP 	180
45	orf20.pep	YFDP PVTAXAWAVFVG GILQLXFQLPWLAKLGF LKLPKLSFKDAAVNRVMKQMAPAILGV 	240
	orf20ng	YFDP PVTALAWAVFVG GILQLGFQLPWLAKLGF LKLPKLNFKDAAVNRVMKQMAPAILGV 	240
	orf20.pep	SVAQVSLVINTIFASYLQSGSVSWMYYADRMMELPSGVLGAALGTILLPTLSKHSANQDT 	300
50	orf20ng	SVAQISLVINTIFASYLQSGSVSWMYYADRMMELPGGVLGAALGTILLPTLSKHSANQDT 	300
	orf20.pep	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHAIAYSFG 	360
55	orf20ng	EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG 	360
	orf20.pep	LIGLIMIKVLAPGFYARQNIXXPVKIAIFTLICXQLMNLXFXGPLXXIGLSLAIGLGACI 	420
	orf20ng	LIGLIMIKVLASGFYARQNIKTVPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI 	420
60	orf20.pep	NAGLLFYLLRRHGIYQPXQGLGSVLXQKCCSRSP 454 	
	orf20ng	NAGLLFFLFRKHGIYRPGQGLGQPSWRKCCSRSP 454 	

An ORF20ng nucleotide sequence <SEQ ID 119> was predicted to encode a protein having amino acid sequence <SEQ ID 120>;

1 MNMLGALAKV GSLTMVSRVL GFVRDTVIAF AFGAGMATDA FFVAFKLPNL  
 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLIVVTA  
 101 LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVGS  
 151 LNSYHKFGIP AFTPTFLNIS FIVFALFFVP YFDPPTVTLA WAVEVGGILQ  
 201 LGFQLPWLAK LGFLKLPKLN FKDAAVNRVM KQMAPAILGV SVAQISLVIN  
 251 TIFASYLQSG SVSWMYADR MMELPGGVLG AALGTILLPT LSKHSANQDT  
 301 EQFSALLDWG LRLCMLTLP AAAGLAVLSF PLVATLFMYR EFTLFDAQMT  
 351 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNL  
 401 FIGPLKHAGL SLAIGLGACI NAGLLFFLFR KHGIYRPGQG LGQPSWRKCC  
 451 SRSP\*

Further DNA sequence analysis revealed the following DNA sequence <SEQ ID 121>:

1 ATGAATATGC TTGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTG  
 51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCGGCG  
 101 CGGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAAACCT GCCCAACCTG  
 151 CTTGCGCGCG TGTTTGCAGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT  
 201 TTTGGCGGAA TATAAGGAAA CGCGTCTTAA AGAGGCGAag gAGGCTTTTA  
 251 TCCGCCACGt tgcgggAatg CTGTCGTTT TGCTGATcgt cGttacCGCG  
 301 CTGGGCATAC TTGCCCGcgc tTGGGTGATT TATGTTtccg CgcccGGCTT  
 351 TACCAAAGAC GCGGACAAGT TCCAACCTTC CATCAGCCTG CTGCGGATTA  
 401 CGTTTCCTTA TATATTATG ATTTCTTTGT CTCTTTTGT CGGCTCGATA  
 451 CTCAATTCCT ACCATAAGTT CGGCATTCCC GCGTTTACGC CCACGTTTTT  
 501 AAACATCTCT TTTATCGTAT TCGCACTGTT TTTCTGCGCG TATTTCGATC  
 551 CGCCCGTTAC CGCGCTGGCG TGGGCGGTTT TTGTCGGCGG TATTTTGCAG  
 601 CTCGGTTTCC AACTGCCGTG GCTGGCGAAA CTGGGCTTTT TGAAACTGCC  
 651 CAAACTGAAT TTCAAAGATG CGGCGGTCAA CCGCGTCATG AAACAGATGG  
 701 CGCCTGCGAT TTTGGGCGTG agcgTGGCGC AAATTTCTTT GgttATCAAC  
 751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTatta  
 801 cgCCGACCGC ATGATGGAGc tgcgcccGGG CGTGCTGGGG GCTGCACTCG  
 851 GTACAATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAA CCAAGATACG  
 901 GAACAGTTT CCGCCCTGCT CGACTGGGGT TTGCGCCTGT GCATGCTGCT  
 951 GACGCTGCCG GCGGCGGcgc GACTGGCGGT ATTGTCGTT CCGCTGGTGG  
 1001 CGACGCTGTT TATGTACCGA GAATTCACGC TGTTTGACGC ACAATGACG  
 1051 CAACACGCGC TGATTGCCCTA TTCTTTCCGT TTAATCGGT TAATTATGAT  
 1101 TAAAGTGTG GCATCCGGCT TTTATGCGCG GCAAAACATC AAAACGCCCG  
 1151 TCAAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTCGCC  
 1201 TTTATCGGTC CGTTGAAACA CGCCGGGCTT TCGCTCGCCA TCGGCCTGGG  
 1251 CGCGTGCATC AACGCCGGAT TGTTGTCTT CCTGTTGCGC AAACACGGTA  
 1301 TTTACCGGCC cggcaggggt tgggcggcgt TCTTGGCGAA AATGCTGCTC  
 1351 GCGCTCGCG TGATGTGCGG CGGACTGTGG GCGGCGCAGG CTTGCTGCGC  
 1401 GTTCGAATGG GCGCACGCCG GCGGAATGCG GAAAGCGGG CAGCTCTGCA  
 1451 TCCTGATTGC CGTCGGCGGC GGAAGTATT TCGCATCTCT GCGGCTTTG  
 1501 GGCTTCCGTC CGGCCATTT CAAACGCGTG GAAAGCTGA

This encodes the following amino acid sequence <SEQ ID 122; ORF20ng-1>:

1 MNMLGALAKV GSLTMVSRVL GFVRDTVIAF AFGAGMATDA FFVAFKLPNL  
 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLIVVTA  
 101 LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVGS  
 151 LNSYHKFGIP AFTPTFLNIS FIVFALFFVP YFDPPTVTLA WAVEVGGILQ  
 201 LGFQLPWLAK LGFLKLPKLN FKDAAVNRVM KQMAPAILGV SVAQISLVIN  
 251 TIFASYLQSG SVSWMYADR MMELRRGVLG AALGTILLPT LSKHSANQDT  
 301 EQFSALLDWG LRLCMLTLP AAAGLAVLSF PLVATLFMYR EFTLFDAQMT  
 351 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNL  
 401 FIGPLKHAGL SLAIGLGACI NAGLLFFLLR KHGIYRPGRG WAFLAKMLL  
 451 ALAVMCGGLW AAQACLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAAL  
 501 GFRPRHFKRV ES\*

ORF20ng-1 and ORF20-1 show 95.7% identity in 512 aa overlap:

	10	20	30	40	50	60
orf20-1.pep	MNMLGALAKV	GSLTMVSRVL	GFVRDTVIAF	AFGAGMATDA	FFVAFKLPNL	LRRVFAEGAF
orf20ng-1	MNMLGALAKV	GSLTMVSRVL	GFVRDTVIAF	AFGAGMATDA	FFVAFKLPNL	LRRVFAEGAF
	70	80	90	100	110	120
orf20-1.pep	AQAFVPILAE	YKETRSKEA	EAFIRHVAG	MLSFVLIV	TALGILAAP	WVIYVSAP
orf20ng-1	AQAFVPILAE	YKETRSKEA	EAFIRHVAG	MLSFVLIV	TALGILAAP	WVIYVSAP

-121-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf20-1.pep	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFPTPTFLNVSFIVFALFFVP					
	orf20ng-1	ADKFQLSISLRLRITFPYILLISLSSFVGSILNSYHKFGIPAFPTPTFLNISFIVFALFFVP					
		130	140	150	160	170	180
10	orf20-1.pep	YFDPPTALAWAVFVGILQLGFQPLWAKLGLKLPKLSFKDAAVNRVMQMAPAILGV					
	orf20ng-1	YFDPPTALAWAVFVGILQLGFQPLWAKLGLKLPKLNFKDAAVNRVMQMAPAILGV					
		190	200	210	220	230	240
15	orf20-1.pep	SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT					
	orf20ng-1	SVAQISLVINTIFASYLQSGSVSWMYADRMELRRGVLGAALGTILLPTLSKHSANQDT					
		250	260	270	280	290	300
20	orf20-1.pep	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG					
	orf20ng-1	EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG					
25		310	320	330	340	350	360
30	orf20-1.pep	LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI					
	orf20ng-1	LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI					
		370	380	390	400	410	420
35	orf20-1.pep	NAGLLFYLLRRHGIYQPGKWAAFLAKMLLSLAVMCGGLWAAQAYLPFEWAHAGGMRKAG					
	orf20ng-1	NAGLLFLLRKHGIYRPGRWAAFLAKMLLALAVMCGGLWAAQACLPFEWAHAGGMRKAG					
		430	440	450	460	470	480
40	orf20-1.pep	QLCILIAVGGGLYFASLAALGFRPRHFKRVENX					
	orf20ng-1	QLCILIAVGGGLYFASLAALGFRPRHFKRVESX					
		490	500	510			

In addition, ORF20ng-1 shows significant homology with a virulence factor of *S.typhimurium*:

45	sp P37169 MVIN SALTY VIRULENCE FACTOR MVIN pir  S40271 mviN protein - Salmonella typhimurium gi 438252 (226133) mviB gene product [Salmonella typhimurium] gnl PID d1005521 (D25292) ORF2 [Salmonella typhimurium] Length = 524 Score = 1573 (750.1 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220 Identities = 309/467 (66%), Positives = 368/467 (78%)
50	Query: 1 MNMLGALAKVGSILTMVSRVLGFVRDVTIARAFAFGMATDAFFVAFKLPNLLRRVFAEGAF 60 MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDAFFVAFKLPNLLRR+FAEGAF Sbjct: 14 MNLKSLAAVSSMTMFSRVLGFARDAIVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF 73
55	Query: 61 AQAFVPILAEYKETSKEATEAFIRHVAGMLSFVLIVVTALGILAAPWVIYVSAPGFTKD 120 +QAQFVPILAEYK + +EAT F+ +V+G+L+ L VVT G+LAAPWVI V+APGF Sbjct: 74 SQAQFVPILAEYKSKQGEETRIFVAYVSGLLTLALAVVTVAGMLAAPWVIMVTAPGFADT 133
60	Query: 121 ADKFQLSISLRLRITFPYILLISLSSFVGSILNSYHKFGIPAFPTPTFLNISFIVFALFFVP 180 ADKF L+ LLRITFPYILLISL+S VG+ILN+++F IPAF PTFLNIS I FALF P Sbjct: 134 ADKFALTTQLLRITFPYILLISLASLVGAILNTWNRSIPAFAPTFLNISMIGFALFAAP 193
65	Query: 181 YFDPPTALAWAVFVGILQLGFQPLWAKLGLKLPKLNFKDAAVNRVMQMAPAILGV 240 YF+PPV ALAWAV VGG+LQL +QLP+L K+G L LP++NF+D RV+QOM PAILGV Sbjct: 194 YFNPPVLALAWAVTVGGVLQLVYQLPYLKKIGMLVLPRIINFRDTGAMRVVQMGPAAILGV 253
70	Query: 241 SVAQISLVINTIFASYLQSGSVSWMYADRMELRRGVLGAALGTILLPTLSKHSANQDT 300 SV+QISL+INTIFAS+L SGSVSWMYADR+ME GVLG ALGTILLP+LSK A+ + Sbjct: 254 SVSQISLIINTIFASFLASGSVSWMYADRMEFPSSGVLGVALGTILLPSLSKSFASGNH 313

-122-

Query: 301 EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLEMYREFTLFDAQMTQHAIAYSFG 360  
 +++ L+DWGLRLC LL LP+A L +L+ PL +LF Y +FT FDA MTQ ALIAYS G  
 Sbjct: 314 DEYCRMLDWGLRLCFLALPSAVALGILAKPLTVSLFQYGKFTAFDAAMTQRALIAYSFG 373

5 Query: 361 LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLACI 420  
 LIGLI++KVLA GFY+RQ+IKTPVKIAI TLI TQLMNLAFIGPLKHAGLSL+IGL AC+  
 Sbjct: 374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433

10 Query: 421 NAGLLFFLLRKHGIYRPGRGWXXXXXXXXXXXXXVMCGGLWAAQACLP 467  
 NA LL++ LRK I+ P GW VM L+ +P  
 Sbjct: 434 NASLLYWQLRKQNIFTPQPGWMWFLMRIISVLVMAAVLFGVLHIMP 480

Score = 70 (33.4 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220  
 Identities = 14/41 (34%), Positives = 23/41 (56%)

15 Query: 469 EWAHAGGMRKAGQLCILIAVGGGLYFASLAALGFRPRHFKR 509  
 EW+ + + +L ++ G YFA+LA LGF+ + F R  
 Sbjct: 481 EWSQGSMLWRLRLMAVVIAGIAAYFAALAVLGFVKVEFVR 521

- 20 Based on this analysis, including the homology with a virulence factor from *S.typhimurium*, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 15

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 123>:

25 1 atGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA  
 51 GCAAGCCGTT TACGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG  
 101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC  
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT  
 201 GTTTACTGCG CCGGCTTCAG GCAAAATCGC CGCGATTAC CGTGGCGAAA  
 30 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC  
 301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACTTAA GCGGCGAAGA  
 351 AGTGCGCCGC AACCTGATCC AATCCGTTT GTGGACTGCG CTGCGCACCC  
 401 GTCCGTTTCA CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC  
 451 GTCAATGCGA TGGACACCAA TCCG..

- 35 This corresponds to the amino acid sequence <SEQ ID 124; ORF22>:

1 MIKIKKGLNL PIAGRPEQAV YDGPATEVA LLGEEYAGMR PSMKVKEGDA  
 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSV VIAVEXNDEI  
 101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPPFAIF  
 151 VNAMDTNP..

- 40 Further work revealed the complete nucleotide sequence <SEQ ID 125>:

1 ATGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA  
 51 GCAAGCCGTT TACGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG  
 101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC  
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT  
 45 201 GTTTACTGCG CCGGCTTCAG GCAAAATCGC CGCGATTAC CGTGGCGAAA  
 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC  
 301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACTTAA GCGGCGAAGA  
 351 AGTGCGCCGC AACCTGATCC AATCCGTTT GTGGACTGCG CTGCGCACCC  
 401 GTCCGTTTCA CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC  
 50 451 GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTATTAT  
 501 CAAAGAAGCC GCCGAGGATT TCAAACGCGG CCTGTTGGTA TTGAGCCGTT  
 551 TGACCGAACG CAAAATCCAT GTTTGTAAGG CAGCTGGCGC AGACGTGCCG  
 601 TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC  
 651 TGCCGTTTGG AGTGGCACGC ACATTCAATT CATCGAGCCG GTCGCGCGGA  
 701 ATAAAACCGT GTGGACCATC AATTATCAAG ATGTAATTAC CATTGGCCGT  
 751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CCCTAGGTGG  
 801 TTCTCAAGTC AACAAACCGC GCCTCTTGCG TACCGTTTGG GGTGCGAAG  
 851 TATCGCAAAT TACTCGGGC GAATTGGTTG ACACAGACAA CCGCGTGATT  
 901 TCCGTTTCGG TATTGAACGG CGCGATTACA CAAGCGCGCG ACGATTATTT

-123-

```

951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
1001 AGCTGTTTCGG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT
1051 ACAACCCTCG GCCATTTTCCT GAAAAACAAA CTCTTCAAGT TCAACACAGC
1101 CGTCAACGGC GGCACCCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG
5 1151 TGATGCCCTT GGATATCCTG CCCACCCTGC TTTTGC GCGA TTTAATCGTC
1201 GGCATACCG ACAGCGCGCA GGCATTGGGT TGCTTGAAT TGGACGAAGA
1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC
1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA

```

This corresponds to the amino acid sequence <SEQ ID 126; ORF22-1>:

```

10 1 MIKIKKGLNL PIAGRPEQAV YDGPATEVA LLGEEYAGMR PSMKVKEGDA
51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSV VIAVEGNDEI
101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADV
201 SENAANIETH EFGGPHFAGL SGTHIHFIET VGANKTVWTI NYQDVITIGR
15 251 LFATGRLNTE RVIALGGSQV NKPRLLRTVL GAKVSQITAG ELVDTDNRVI
301 SGSVLNGAIT QGAHDYLGRI HNQISVIEEG RSKELFGWVA PQPDKYSITR
351 TTLGHFLKNK LFKFNTAVNG GDRAMVPIGT YERVMPLDIL PTLRLDLIV
401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRKVL ETIEKEG*

```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 127>:

```

20 1 ATGATTAAAA TCAAAAAGG TCTAACCTG CCCATCGCGG GCAGACCGGA
51 GCAAGTCATT TATGACGGGC CCGTCATTAC CGAAGTCGCG TTGCTTGGCG
101 AAGAATATGC CGGTATGCGC CCCTNGATGA AAGTCAAGGA AGGCGATGCC
151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGNATC CGGGCGTGGT
201 GTTACCGCG CCNGTTTCAG GCAAATCGC CGCCATCCAT CGCGGCGAAA
25 251 AGCGCGTACT TCAGTCGGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
301 GAGTTTGAAC GCTACGCGCC CGAAGCGTTG GCAAACCTAA GCGGCGGANGA
351 ANTNGNNGC AATCTGATCC AATCCGGTTT GTGGACTGCG CTGCGTANCC
401 GTCGGTTCAG CAAAATCCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
451 GTCAATGCGA TGGACACCAA TCCGCTNGCG GCAGACCCTG TGGTTGTGAT
30 501 CAAAGAAGCC GNCGANGATT TCAGACGANG TNTGCTGGTA TTGAGCCGTT
551 TGACCGAGCG TAAAATCCAT GTGTGTAAGG CAGCTGGCGC AGACGTGCCG
601 TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC
651 GGCGGGTTTG AGTGGCACGC ACATTCATTT CATTGAGCCG GTCGGTGCAA
701 ACAAACCGT TTGGACCATC AATTATCAAG ATGTAATTGC CATCGGACGT
35 751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CTTTGGGTGG
801 TTCTCAAGTC AACAAACCAC GCCTCTTGCG TACCGTTTGT GGTGCGAAAG
851 TATCGCAAAT TACTGCGGGC GAATTGGTTG ACGCAGACAA CCGCGTGATT
901 TCCGGTTCG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT
951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
40 1001 AGCTGTTTCGG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT
1051 ACGACCCTCG GCCATTTTCCT GAAAAACAAA CTCTTCAAGT TCACGACAGC
1101 CGTCAACGGT GGCACCCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG
1151 TAATGCCGCT AGACATCCTG CCTACCCTGC TTTTGC GCGA TTTAATCGTC
1201 GGCATACCG ACAGCGCGCA AGCATTTGGT TGCTTGAAT TGGACGAAGA
45 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATANGGCC
1301 CGCTGTTGCG TAAGGTGCTG GAAACCNTG AGAAGGAAGG CTGA

```

This encodes a protein having amino acid sequence <SEQ ID 128; ORF22a>:

```

50 1 MIKIKKGLNL PIAGRPEQVI YDGPVITEVA LLGEEYAGMR PSMKVKEGDA
51 VKKGQVLFED KXPGVVFTA PVSGKIAAIH RGEKRVLSV VIAVEGNDEI
101 EFERYAPEAL ANLSGXEXX NLIQSGLWTA LRXPFSKIP AVDAEPFAIF
151 VNAMDTNPLA ADPVVVIKEA XDFRXXLV LSRLTERKIH VCKAAGADV
201 SENAANIETH EFGGPHFAGL SGTHIHFIET VGANKTVWTI NYQDVIAIGR
251 LFATGRLNTE RVIALGGSQV NKPRLLRTVL GAKVSQITAG ELVDADNRVI
301 SGSVLNGAIT QGAHDYLGRI HNQISVIEEG RSKELFGWVA PQPDKYSITR
55 351 TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMPLDIL PTLRLDLIV
401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EXGPLLRKVL ETXEKEG*

```

The originally-identified partial strain B sequence (ORF22) shows 94.2% identity over a 158aa overlap with ORF22a:

```

60 orf22.pep      10      20      30      40      50      60
      MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED
orf22a  MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED

```

-121-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf20-1.pep	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFTPTFLNVSFIVFALFFVP					
	orf20ng-1	ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFTPTFLNISFIVFALFFVP					
		130	140	150	160	170	180
10	orf20-1.pep	YFDPPTALAWAVFVGILQLGFQPLWAKLGLKLPKLSFKDAAVNRVMQMAPAILGV					
	orf20ng-1	YFDPPTALAWAVFVGILQLGFQPLWAKLGLKLPKLNFKDAAVNRVMQMAPAILGV					
		190	200	210	220	230	240
15	orf20-1.pep	SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT					
	orf20ng-1	SVAQISLVINTIFASYLQSGSVSWMYADRMELRRGVLGAALGTILLPTLSKHSANQDT					
		250	260	270	280	290	300
20	orf20-1.pep	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG					
	orf20ng-1	EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG					
25		310	320	330	340	350	360
30	orf20-1.pep	LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI					
	orf20ng-1	LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI					
		370	380	390	400	410	420
35	orf20-1.pep	NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMCGGLWAAQAYLPFEWAHAGGMRKAG					
	orf20ng-1	NAGLLFLLRKHGIYRPRGWAAFLAKMLLALAVMCGGLWAAQACLPFEWAHAGGMRKAG					
		430	440	450	460	470	480
40	orf20-1.pep	QLCILIAVGGGLYFASLAALGFRPRHFKRVENX					
	orf20ng-1	QLCILIAVGGGLYFASLAALGFRPRHFKRVESX					
		490	500	510			

In addition, ORF20ng-1 shows significant homology with a virulence factor of *S.typhimurium*:

45	sp P37169 MVIN SALTY VIRULENCE FACTOR MVIN pir  S40271 mviN protein - Salmonella typhimurium gi 438252 (226133) mviB gene product [Salmonella typhimurium] gnl PID d1005521 (D25292) ORF2 [Salmonella typhimurium] Length = 524 Score = 1573 (750.1 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220 Identities = 309/467 (66%), Positives = 368/467 (78%)
50	Query: 1 MNMLGALAKVGSILTMVSRVLGFVRDVTIARAFAFGAGMATDAFFVAFKLPNLLRRVFAEGAF 60 MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDAFFVAFKLPNLLRR+FAEGAF Sbjct: 14 MNLKSLAAVSSMTMFSRVLGFARDAIVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF 73
55	Query: 61 AQAFVPILAEYKETSKEATEAFIRHVAGMLSFVLIVVTALGILAAPWVIYVSAPGFTKD 120 +QAQFVPILAEYK + +EAT F+ +V+G+L+ L VVT G+LAAPWVI V+APGF Sbjct: 74 SQAFVPILAEYKSKQGEETRIFVAYVSGLLTLALAVVTVAGMLAAPWVIMVTAPGFADT 133
60	Query: 121 ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFTPTFLNISFIVFALFFVP 180 ADKF L+ LLRITFPYILLISL+S VG+ILN+++F IPAF PTFLNIS I FALF P Sbjct: 134 ADKFALTTQLLRITFPYILLISLASLVGAILNTWNRFSIPAFAPTFLNISMIGFALFAAP 193
65	Query: 181 YFDPPTALAWAVFVGILQLGFQPLWAKLGLKLPKLNFKDAAVNRVMQMAPAILGV 240 YF+PPV ALAWAV VGG+LQL +QLP+L K+G L LP++NF+D RV+KQM PAIGV Sbjct: 194 YFNPPVLALAWAVTVGGVQLVYQLPYLKKIGMLVLPINFRDGTAMRVVQMGPAIGV 253
70	Query: 241 SVAQISLVINTIFASYLQSGSVSWMYADRMELRRGVLGAALGTILLPTLSKHSANQDT 300 SV+QISL+INTIFAS+L SGSVSWMYADR+ME GVLG ALGTILLP+LSK A+ + Sbjct: 254 SVSQISLIINTIFASFLASGSVSWMYADRLMEFPGSVLGVALGTILLPSLSKSFASGNH 313

Query: 301 EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG 360  
 +++ L+DWGLRLC LL LP+A L +L+ PL +LF Y +FT FDA MTQ ALIAYS G  
 Sbjct: 314 DEYCRMLMDWGLRLCFLALPSAVALGILAKPLTVSLFYQYGFATFADAAMTQRALIAYSVG 373

5 Query: 361 LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI 420  
 LIGLI++KVLG GFY+RQ+IKTPVKIAI TLI TQLMNLAFIGPLKHAGLSL+IGL AC+  
 Sbjct: 374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433

10 Query: 421 NAGLLFFLLRKHGIIYRPGRGWXXXXXXXXXXXXXVMCGGLWAAQACLP 467  
 NA LL++ LRK I+ P GW VM L+ +P  
 Sbjct: 434 NASLLYWQLRKQNIPTPQPGWMWFLMRLIISVLVMAAVLFGVLHIMP 480

Score = 70 (33.4 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220  
 Identities = 14/41 (34%), Positives = 23/41 (56%)

15 Query: 469 EWAHAGGMRKAGQLCILIAVGGGLYFASLAALGFRPRHFKR 509  
 EW+ + + +L ++ G YFA+LA LGF+ + F R  
 Sbjct: 481 EWSQGSMLWRLRLMAVVIAGIAAYFAALAVLGFKVKEFVR 521

- 20 Based on this analysis, including the homology with a virulence factor from *S.typhimurium*, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 15

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 123>:

25 1 atGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA  
 51 GCAAGCCGTT tACGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG  
 101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC  
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT  
 201 GTTTACTGCG CCGGCTTCAG GcAAAATCGC CGCGATTAC CGTGGCGAAA  
 30 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC  
 301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACCTAA GCGGCGAAGA  
 351 AGTGCGCCGC AACCTGATCC AATCCGTTT GTGGACTGCG CTGCGCACCC  
 401 GTCCGTTTCA CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC  
 451 GTCATGCGA tGGACACCAA TCCG..

- 35 This corresponds to the amino acid sequence <SEQ ID 124; ORF22>:

1 MIKIKKGLNL PIAGRPEQAV YDGPAITEVA LLGEEYAGMR PSMKVKEGDA  
 51 VKKGQVLFED KKNPGVVETA PASGKIAAIH RGEKRVLQSV VIAVEXNDEI  
 101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF  
 151 VNAMDTP..

- 40 Further work revealed the complete nucleotide sequence <SEQ ID 125>:

1 ATGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA  
 51 GCAAGCCGTT TACGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG  
 101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC  
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT  
 45 201 GTTTACTGCG CCGGCTTCAG GcAAAATCGC CGCGATTAC CGTGGCGAAA  
 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC  
 301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACCTAA GCGGCGAAGA  
 351 AGTGCGCCGC AACCTGATCC AATCCGTTT GTGGACTGCG CTGCGCACCC  
 401 GTCCGTTTCA CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC  
 50 451 GTCATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTCAATTAT  
 501 CAAAGAAGCC GCCGAGGATT TCAAACGCGG CCTGTTGGTA TTGAGCCGTT  
 551 TGACCGAAGC CAAAATTCAT GTTTGTAAGG CAGCTGGCGC AGACGTGCCG  
 601 TCTGAAAATG CTGCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC  
 651 TGCCGGTTTG AGTGGCACGC ACATTCATTT CATCGAGCCG GTCGGCGCGA  
 701 ATAAACCGT GTGGACCATC AATTATCAAG ATGTAATTAC CATTGGCCGT  
 751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CCCTAGGTGG  
 801 TTCTCAAGTC AACAAACCGC GCCTCTTGGC TACCGTTTGG GGTGCGAAAG  
 851 TATCGCAAAT TACTGCGGGC GAATTGGTTG ACACAGACAA CCGCGTGATT  
 901 TCCGGTTTCG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT



-123-

5  
 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG  
 1001 AGCTGTTTCGG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT  
 1051 ACAACCCTCG GCCATTTTCCT GAAAAACAAA CTCTTCAAGT TCAACACAGC  
 1101 CGTCAACGGC GGCAGCCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG  
 1151 TGATGCCCTT GGATATCCTG CCCACCCTGC TTTTGC GCGA TTTAATCGTC  
 1201 GCGGATACCG ACAGCGCGCA GGCATTGGGT TGCTTGAAT TGGACGAAGA  
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC  
 1301 CGCTGTTGCG CAAAGTGCTG GAAACCATG AGAAGGAAGG CTGA

This corresponds to the amino acid sequence <SEQ ID 126; ORF22-1>:

10  
 1 MIKIKKGLNL PIAGRPEQAV YDGPATEVA LLGEEYAGMR PSMKVKEGDA  
 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSV VIAVEGNDEI  
 101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF  
 151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADVP  
 201 SENAANIETH EFGGPHFAGL SGTHIHFIET VGANKTVWTI NYQDVITIGR  
 15  
 251 LFATGRLNTE RVIALGGSQV NKPRLLRTVL GAKVSQITAG ELVDTDNRVI  
 301 SGSVLNGAIT QGAHDYLGRI HNQISVIEEG RSKELFGWVA PQPKYSITR  
 351 TTLGHFLKNK LFKFNTAVNG GDRAMVPIGT YERVMPLDIL PTLRLDLIV  
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRKVL ETIEKEG\*

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 127>:

20  
 1 ATGATTAAAA TCAAAAAGG TCTAACCTG CCCATCGCGG GCAGACCGGA  
 51 GCAAGTCATT TATGACGGGC CCGTCATTAC CGAAGTCGCG TTGCTTGGCG  
 101 AAGAATATGC CGGTATGCGC CCCTNGATGA AAGTCAAGGA AGGCGATGCC  
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAAGNATC CGGGCGTGCT  
 201 GTTTACCGCG CCNGTTTCAG GCAAAATCGC CGCCATCCAT CGCGCGGAAA  
 25  
 251 AGCGCGTACT TCAGTCGGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC  
 301 GAGTTCTGAAC GCTACGCGCC CGAAGCGTTG GCAAACCTAA GCGGCGGANGA  
 351 ANTNNNGNGC AATCTGATCC AATCCGGTTT GTGGACTGCG CTGCGTANCC  
 401 GTCGGTTCAG CAAAATCCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC  
 451 GTCAATGCGA TGGACACCAA TCCGCTNGCG GCAGACCCTG TGGTTGTGAT  
 30  
 501 CAAAGAAGCC GNCGANGATT TCAGACGANG TNTGCTGGTA TTGAGCCGTT  
 551 TGACCGAGCG TAAAATCCAT GTGTGTAAGG CAGCTGGCGC AGACGTGCCG  
 601 TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC  
 651 GGCCGGTTTG AGTGGCACGC ACATTCAATT CATTGAGCCG GTCGGTGCAA  
 701 ACAAACCGT TTGGACCATC AATTATCAAG ATGTAATTGC CATCGGACGT  
 35  
 751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CTTTGGGTGG  
 801 TTCTCAAGTC AACAAACCAC GCCTCTTGCG TACCGTTTTG GGTGCGAAAG  
 851 TATCGCAAAT TACTGCGGGC GAATTGGTTG ACGCAGACAA CCGCGTGATT  
 901 TCCGGTTCG TATTGAACGG CGCGATTACA CAAGCGCGCG ACGATTATTT  
 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG  
 40  
 1001 AGCTGTTTCGG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT  
 1051 ACGACCCTCG GCCATTTTCCT GAAAAACAAA CTCTTCAAGT TCACGACAGC  
 1101 CGTCAACGGT GGCAGCCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG  
 1151 TAATGCCGCT AGACATCCTG CCTACCCTGC TTTTGC GCGA TTTAATCGTC  
 1201 GCGGATACCG ACAGCGCGCA AGCATTGGGT TGCTTGAAT TGGACGAAGA  
 45  
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATANGGCC  
 1301 CGCTGTTGCG TAAGGTGCTG GAAACNTTG AGAAGGAAGG CTGA

This encodes a protein having amino acid sequence <SEQ ID 128; ORF22a>:

50  
 1 MIKIKKGLNL PIAGRPEQVI YDGPVITEVA LLGEEYAGMR PSMKVKEGDA  
 51 VKKGQVLFED KXPGVVFTA PVSIGKIAAIH RGEKRVLSV VIAVEGNDEI  
 101 EFERYAPEAL ANLSGXEXXX NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF  
 151 VNAMDTNPLA ADPVVVIKEA XXDFRXXLV LSRLTERKIH VCKAAGADVP  
 201 SENAANIETH EFGGPHFAGL SGTHIHFIET VGANKTVWTI NYQDVIAIGR  
 251 LFATGRLNTE RVIALGGSQV NKPRLLRTVL GAKVSQITAG ELVDADNRVI  
 301 SGSVLNGAIT QGAHDYLGRI HNQISVIEEG RSKELFGWVA PQPKYSITR  
 55  
 351 TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMPLDIL PTLRLDLIV  
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EXGPLLRKVL ETXEKEG\*

The originally-identified partial strain B sequence (ORF22) shows 94.2% identity over a 158aa overlap with ORF22a:

60  
 orf22.pep MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED  
 orf22a MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMKVKEGDAVKKGQVLFED

-124-

		10	20	30	40	50	60
		70	80	90	100	110	120
5	orf22.pep	KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGEEVRR					
	orf22a	KKXPGVVFTAPVSGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGXEXXX					
		70	80	90	100	110	120
10		130	140	150			
	orf22.pep	NLIQSGLWTALRTRPFSKIPAVDAEPPFAIFVNAMDTNP					
	orf22a	NLIQSGLWTALRXRPFSKIPAVDAEPPFAIFVNAMDTNP LAADPVVVIKEAXXDFRXXLV					
		130	140	150	160	170	180

The complete strain B sequence (ORF22-1) and ORF22a show 94.9% identity in 447 aa overlap:

15	orf22a.pep	10	20	30	40	50	60
	orf22-1	MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMKVKEGDAVKKGQVLFED      :     :     :     :     :     :     :     :     :     :					
20	orf22a.pep	70	80	90	100	110	120
	orf22-1	KKXPGVVFTAPVSGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGXEXXX         :     :     :     :     :     :     :     :					
25	orf22a.pep	130	140	150	160	170	180
	orf22-1	NLIQSGLWTALRXPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFFRXXLV      :     :     :     :     :     :     :     :     :					
30	orf22a.pep	190	200	210	220	230	240
	orf22-1	LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFIIEPVGANKTVWTI      :     :     :     :     :     :     :     :     :					
35	orf22a.pep	250	260	270	280	290	300
	orf22-1	NYQDVIAIGRLFATGRLNTERVIALGGSQVKNPRLRLTVLGAKVSQITAGELVDADNRVI      :     :     :     :     :     :     :     :     :					
40	orf22a.pep	310	320	330	340	350	360
	orf22-1	SGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPDKYSITRTTLGHFLKNK      :     :     :     :     :     :     :     :     :					
45	orf22a.pep	370	380	390	400	410	420
	orf22-1	LFKFTTAVNGGDRAMVPIGTYERVMPLDILPTLLLRDLIVGDTDSAQALGCLELDEEDLA     :     :     :     :     :     :     :     :     :					
50	orf22a.pep	430	440				
	orf22-1	LCSFVCPGKYEXGPLLRKVLETXEKEGX      :     :     :     :     :     :     :     :     :					
55	orf22a.pep	430	440				
	orf22-1	LCSFVCPGKYEYGPLLRKVLETIEKEGX      :     :     :     :     :     :     :     :     :					

Further work identified a partial gene sequence <SEQ ID 129> from *N.gonorrhoeae*, which encodes the following amino acid sequence <SEQ ID 130; ORF22ng>:

65	1	MIKIKKGLNL	PIAGRPEQVI	YDGPATEVA	LLGEEYVGM	PSMKIKEGEA
	51	VKKGQVLFED	KKNPGVVFTA	PASGKIAAIH	RGEKRVLSV	VIAVEGNDEI
	101	EFERYVPEAL	AKLSSEKVR	NLIQSGLWTA	LRTRPFSKIP	AVDAEPFAIF

151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADV  
 201 SENAANIETH EFGGPHAGL SGTHIHFI EP VGANKTVWTI NYQDVIAIGR  
 251 LFVTGRLNTE RVVALGGLOV NKPRLLRTVL GAKVSQLTAG ELVDADNRVI  
 301 SGSVLNGAIA QGAHDYLG RY HN\*

# 5 Further work identified complete gonococcal gene <SEQ ID 131>:

1 ATGATTAAAA TCAAAAAAGG TCTAAATCTG CCCATCGCGG GCAGACCGGA  
 51 GCAAGTCATT TATGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG  
 101 AAGAATATGT CGGCATGCGC CCCTCGATGA AAATCAAGGA AGGTGAAGCC  
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTAGT  
 10 201 ATTTACTGCG CCGGCTTCAG GCAAAATCGC CGCTATTAC CGTGGCGAAA  
 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC  
 301 GAGTTCGAAC GCTACGTACC TGAAGCGCTG GCAAAATTGA GCAGCGAAAA  
 351 AGTGCGCCGC AACCTGATTC AATCAGGCTT ATGGACTGCG CTTGCGACCC  
 401 GTCCGTTTCAG CAAAATCCCT GCCGTAGATG CCGAGCCGTT CGCCATCTTC  
 15 451 GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTTCATCAT  
 501 CAAAGAAGCC GCCGAAGACT TCAAACGCGG CCTGTTGGTA TTGAGCCGCC  
 551 TGACCGAAGC TAAAATCCAT GTGTGTAAG CAGCAGCGCG AGACGTGCCG  
 601 TCTGAAAATG CTGCCAATAT CGAAACACAT GAATTTGGCG GCCCGCATCC  
 651 TGCCGGCTTG AGTGGCACGC ACATTCATTT CATCGAGCCA GTCCGCGCGA  
 20 701 ATAAACCCTG GTGGACCATC AATTATCAAG ACGTGATTGC TATCGGACGT  
 751 TTGTTTCGTAA CAGGCCGTCT GAATACCGAG CGCGTGGTTG CCTTGGCGCG  
 801 CCTGCAAGTC AACAAACCGC GCCTCTTGCG TACCGTTTGG GGTGCGAAGG  
 851 TGTCTCAACT TACCGCCGGC GAATTTGGTTG ACGCGGACAA CCGCGTGATT  
 901 TCCGGTTTCG TATTGAACGG TGCGATTGCA CAAGGCGCGC ATGATTATTT  
 25 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG  
 1001 AGCTGTTTCG CTGGGTTGCG CCGCAGCCGG ACAAACTACT CATCACGCGC  
 1051 ACCACTCTCG GCCATTTCTT AAAAAACAAA CTCTTCAAGT TCACGACAGC  
 1101 CGTCAACGGC GCGGACCGCG CCATGGTACC GATCGGCACT TATGAGCGCG  
 1151 TAATGCCGTT GGACATCCTG OCTACCTTGC TTTTGC GCGA TTTAATCGTC  
 30 1201 GCGGATACCG ACAGCGCGCA GGCTTTGGGT TGCTTGAAT TGGACGAAGA  
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC  
 1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA

This encodes a protein having amino acid sequence <SEQ ID 132; ORF22ng-1>:

1 MIKIKKGLNL PIAGRPEQVI YDGPATEVA LLGEEYVGM RPSMKIKEGEA  
 35 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSV VIAVEGNDEI  
 101 EFERYVPEAL AKLSSEKVR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF  
 151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADV  
 201 SENAANIETH EFGGPHAGL SGTHIHFI EP VGANKTVWTI NYQDVIAIGR  
 40 251 LFVTGRLNTE RVVALGGLOV NKPRLLRTVL GAKVSQLTAG ELVDADNRVI  
 301 SGSVLNGAIA QGAHDYLG RY HNQISVIEEG RSKELEFGWA PQDPKYSITR  
 351 TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV  
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRKVL ETIEKEG\*

The originally-identified partial strain B sequence (ORF22) shows 93.7% identity over a 158aa

# 45 overlap with ORF22ng:

orf22.pep	MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED	60
orf22ng	MIKIKKGLNLPIAGRPEQVIYDGPATEVALLGEEYVGM RPSMKIKEGEAVKKGQVLFED	60
orf22.pep	KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGEEVRR	120
orf22ng	KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYVPEALAKLSSEKVR	120
orf22.pep	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP	158
orf22ng	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV	180

The complete sequences from strain B (ORF22-1) and gonococcus (ORF22ng) show 96.2% identity in 447 aa overlap:

60 orf22-1.pep 10 20 30 40 50 60  
 MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED

		10	20	30	40	50	60
	orf22ng-1	MIKIKKGLNLPIAGRPEQVIYDGPATEVALLGEEYVGMRPSMKIKEGEAVKKQVLFED					
5		70	80	90	100	110	120
	orf22-1.pep	KKNPGVVFETAPASGKIAAIHRGEKRVLSVVIAVEGNDEIEFERYAPEALANLSGEEVRR					
	orf22ng-1	KKNPGVVFETAPASGKIAAIHRGEKRVLSVVIAVEGNDEIEFERYVPEALAKLSSEKVR					
10		70	80	90	100	110	120
	orf22-1.pep	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV					
	orf22ng-1	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV					
15		130	140	150	160	170	180
	orf22-1.pep	LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFIPEVGANKTVWTI					
	orf22ng-1	LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFIPEVGANKTVWTI					
20		190	200	210	220	230	240
	orf22-1.pep	NYQDVITIGRLFATGRLNTERVIALGGSQVKNPRLRLTVLGAKVSQITAGELVDTDNRVI					
	orf22ng-1	NYQDVIAIGRLFVTGRLNTERVVALGGQLQVKNPRLRLTVLGAKVSQITAGELVDADNRVI					
25		250	260	270	280	290	300
	orf22-1.pep	SGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPKYSITRTTLGHFLKNK					
	orf22ng-1	SGSVLNGAIAQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPKYSITRTTLGHFLKNK					
30		310	320	330	340	350	360
	orf22-1.pep	LKFNTAVNGGDRAMVPIGTYERVMPDLPTLLRLDLIVGDTDSAQALGCLELDEEDLA					
	orf22ng-1	LKFNTAVNGGDRAMVPIGTYERVMPDLPTLLRLDLIVGDTDSAQALGCLELDEEDLA					
35		370	380	390	400	410	420
	orf22-1.pep	LCSFVCPGKYEYGPLLRLKVLETIEKEGX					
	orf22ng-1	LCSFVCPGKYEYGPLLRLKVLETIEKEGX					
40		430	440				
	orf22-1.pep	LCSFVCPGKYEYGPLLRLKVLETIEKEGX					
	orf22ng-1	LCSFVCPGKYEYGPLLRLKVLETIEKEGX					
45		430	440				

Homology with 48kDa outer membrane protein of *Actinobacillus pleuropneumoniae* (accession number U24492).

50	Orf22	1	MIIKKKGLNLP MI IKKGL+LPIAG P Q +++G + EVA+LGEEY GMRPSMKV+EGD VKKGQVLFD	60
	48kDa	1	MITIKKGLDLPIAGTPAQVIHNGNTVN MITIKKGLDLPIAGTPAQVIHNGNTVN	60
55	orf22	61	KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI KKNPGVVFTAPASG + I+RGEKRVLSVVI VE +++I F RY LA+LS E+V++	120
	48kDa	61	KKNPGVVFTAPASGTVVTTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ	120
60	orf22	121	NLIQSGLWLTALRTRPFSKIPAVDAEPFAIFVNAMDTNP	158
	48kDa	121	NLI+SGLWTA RTRPFSK+PA+DA P +IFVNAMDTNP NLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNP	158

gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus pleuropneumoniae]  
Length = 449

Score = 530 bits (1351), Expect = e-150

-127-

Identities = 274/450 (60%), Positives = 323/450 (70%), Gaps = 4/450 (0%)

Query: 1 MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMKVKEGDAVKKGQVLFED 60  
 Sbjct: 1 MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGM RPSMKVREGDVVKKGQVLFED 60

Query: 61 KKPVGCVFTAPVSGKIAAIHRGEKRVLSVVI AVEGNDIEFERYAPEALANLSGXEXX 120  
 KK PGVFTAP SG + I+RGEKRVLSVVI VEG+++I F RY LA+LS +  
 Sbjct: 61 KKNPGVFTAPASGTVVVTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

Query: 121 NLIQSGLWTALRXRPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRRXXLV 180  
 NLI+SGLWTA R RPFSSK+PA+DA P +IFVNAMDTNPLAADP VV+KE DF+ V  
 Sbjct: 121 NLIESGLWTAFTTRPFSSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGTLV 180

Query: 181 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHAGLSGTHIHFIIEPVGANKTV 237  
 L+RL ++ +++CK A +++P S I F G HPAGL GTHIHFI++PVGA K V  
 Sbjct: 181 LTRLFNGQKPVYLCCKDADSNIPSPAIIEGITIKSFSGVHPAGLVGTHIHFDVPVGATKQV 240

Query: 238 WTINYQDVIAIGRLFATGRLNTERVIALGGSQVKNPRLRLTVLGAKVSQITAGELVDADN 297  
 W +NYQDVIAIG+LF TG L T+R+I+L G QV PRL+RT LGA +SQ+TA EL +N  
 Sbjct: 241 WHLNYQDVIAIGKLETTGELFTDRIISLAGPQVKNPRLVTRLGANLSQLTANELNAGEN 300

Query: 298 RVISGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPKYSITRTTLGHFL 357  
 RVISGSVL+GA G DYLG RY Q+SV+ EGR KELFGW+ P DK+SITRT LGHF  
 Sbjct: 301 RVISGSVLGATAAGPVDYLG RYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

Query: 358 KNKLFKFTTAVNGGDRAMVPIGTIERVMXXXXXXXXXXXXXXXXXVGD TDSAQXXXXXXXXXX 417  
 K KLF FTTAV+GG+RAMVPIG YERVM GDTDSAQ  
 Sbjct: 361 K-KLFNFTTAVHGERAMVPIGAYERVMPLDIPTLLLRDLAAGDTDSAQNLGCLELDEE 419

Query: 418 XXXXXSFVCPGKYEXGP LLRKVLETXEKEG 447  
 ++VCPGK GP+LR LE EKEG

ORF22ng-1 also shows homology with the OMP from *A. pleuropneumoniae*:

gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus  
 pleuropneumoniae] Length = 449  
 Score = 555 bits (1414), Expect = e-157  
 Identities = 284/450 (63%), Positives = 337/450 (74%), Gaps = 4/450 (0%)

Query: 27 MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYVGM RPSMKIKEGEAVKKGQVLFED 86  
 MI IKKGL+LPIAG P QVI++G + EVA+LGEEYVGM RPSMK++EG+ VKKGQVLFED  
 Sbjct: 1 MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGM RPSMKVREGDVVKKGQVLFED 60

Query: 87 KKNPGVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDIEFERYVPEALAKLSSEKVR 146  
 KKNPGVFTAPASG + I+RGEKRVLSVVI VEG+++I F RY LA LS+E+++  
 Sbjct: 61 KKNPGVFTAPASGTVVVTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

Query: 147 NLIQSGLWTALRTRPFSSKIPAVDAEPFAIFVNAMDTNPLAADPTVVIKEAAEDFKRGLLV 206  
 NLI+SGLWTA RTRPFSSK+PA+DA P +IFVNAMDTNPLAADP V+KE DFK GL V  
 Sbjct: 121 NLIESGLWTAFTTRPFSSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGTLV 180

Query: 207 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHAGLSGTHIHFIIEPVGANKTV 263  
 L+RL ++ +++CK A +++P S I F G HPAGL GTHIHFI++PVGA K V  
 Sbjct: 181 LTRLFNGQKPVYLCCKDADSNIPSPAIIEGITIKSFSGVHPAGLVGTHIHFDVPVGATKQV 240

Query: 264 WTINYQDVIAIGRLFVTGRLNTERVIALGGLQVKNPRLRLTVLGAKVSQITAGELVDADN 323  
 W +NYQDVIAIG+LF TG L T+R+++L G QV PRL+RT LGA +SQLTA EL +N  
 Sbjct: 241 WHLNYQDVIAIGKLETTGELFTDRIISLAGPQVKNPRLVTRLGANLSQLTANELNAGEN 300

Query: 324 RVISGSVLNGAIAQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPKYSITRTTLGHFL 383  
 RVISGSVL+GA A G DYLG RY Q+SV+ EGR KELFGW+ P DK+SITRT LGHF  
 Sbjct: 301 RVISGSVLGATAAGPVDYLG RYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

Query: 384 KNKLFKFTTAVNGGDRAMVPIGTIERVMXXXXXXXXXXXXXXXXXVGD TDSAQXXXXXXXXXX 443  
 K KLF FTTAV+GG+RAMVPIG YERVM GDTDSAQ  
 Sbjct: 361 K-KLFNFTTAVHGERAMVPIGAYERVMPLDIPTLLLRDLAAGDTDSAQNLGCLELDEE 419

Query: 444 XXXXXSFVCPGKYEYGP LLRKVLETIEKEG 473  
 ++VCPGK YGP+LR LE IEKEG  
 Sbjct: 420 DLALCTYVCPGKNNYGPMLRAALEKIEKEG 449

Based on this analysis, including the homology with the outer membrane protein of *Actinobacillus pleuropneumoniae*, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF22-1 (35.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 5A shows the results of affinity purification of the GST-fusion protein, and Figure 5B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 5C). These experiments confirm that ORF22-1 is a surface-exposed protein, and that it is a useful immunogen.

## Example 16

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 133>:

```

1  ..GCGnCGnAAA TCATCCATCC CC..nACGTC GTAGGCCCTG AAGCCAAC TG
51  GTTTTTTATG GTAGCCAGTA CGTTTGTGAT TGCTTTGATT GGTTATTTTG
101 TTACTGAAAA AATCGTCGAA CCGCAATTGG GCCCTTATCA ATCAGATTTG
15  151 TCACAAGAAG AAAAGACAT TCGGCATTCC AATGAAATCA CGCCTTTGGA
201 ATATAAAGGA TTAATTTGGG CTGGCGTGGT GTTTGTTGCC TTATCCGCC
251 TATTGGCTTG GAGCATCGTC CCTGCCGACG GTATTTTGCG TCATCCTGAA
301 ACAGGATTGG TTTCCGGTTC GCCGTTTTTA AAATCGATTG TTGTTTTTAT
351 TTTCTGTGTG TTTGCACTGC CGGGCATTGT TTATGGCCGG GTAACCCGAA
20  401 GTTTGCGCGG CGAACAGGAA GTCGTTAATG CGmyGGCCGA ATCGATGAGT
451 ACTCTGSGGC TTTmTTTGsw CAkCATCTTT TTTGCCGCAC AGTTTGTCGC
501 ATTTTTTAAT TGGACGAATA TTGGGCAATA TATTGCCGTT AAAGGGGCGA
551 CGTTCCTAAA AGAAGTCGGC TTGGGCGGCA GCGTGTGTT TATCGGTTTT
601 ATTTTAATTT GTGCTTTTAT CAATCTGATG ATAGGCTCCG CCTCCGCGCA
25  651 ATGGGCGGTA ACTGCGCCGA TTTTCGTCCC TATGCTGATG TTGGCCGGCT
701 ACGCGCCCGA AGTCAATTCAA GCCGCTTACC GCATCGGTGA TTCCGTTACC
751 AATATTATTA CGCCGATGAT GAGTTATTTT GGGCTGATTA TGGCGACGGT
801 GrkCmmnTAC AAAAAAGATG CGGGCGTGGG TaCGcTGATT wCTATGATGT
851 TGCCGTATTC CGCTTCTTC TTGATTGCgT GGATTGCCTT ATTCTGCATT
30  901 TGGGTATTTg TTTTGGGCCT GCCCGTCGGT CCCGGCGCGC CCACATTCTA
951 TCCCGCACCT TAA

```

This corresponds to the amino acid sequence <SEQ ID 134; ORF12>:

```

1  ..AXXIIHPXXV VGPEANWFFM VASTFVIALI GYFVTEKIVE PQLGPYQSDL
35  51  SQEEKDIRHS NEITPLEYKG LIWAGVVFVA LSALLAWSIV PADGILRHPE
101 TGLVSGSPFL KSIVVFIFLL FALPGIVYGR VTRSLRGEQE VVNAXAESMS
151 TLXLXLXXIF FAAQFVAFFN WTNIGQYIAV KGATFLKEVG LGGSVLFIFG
201 ILICAFINLM IGSASAQWAV TAPIFVPLM LAGYAPEVIQ AAYRIGDSVT
251 NIITPMSYF GLIMATVXXY KKDAGVGTLI XMMLPYSAFF LIAWIALFCI
301 WVFVLGLPVG PGAPTYPAP *

```

Further sequence analysis revealed the complete DNA sequence <SEQ ID 135> to be:

```

1  ATGAGTCAAA CCGATACGCA ACGGGACGGA CGATTTTTAC GCACAGTCGA
51  ATGGCTGGGC AATATGTTGC CGCATCCGGT TACGCTTTTT ATTATTTTCA
101 TTGTGTTATT GCTGATTGCC TCTGCCGTCG GTGCGTATTT CGGACTATCC
151 GTCCCGGATC CGCGCCCTGT TGGTGCGAAA GGACGTGCCG ATGACGGTTT
45  201 GATTTACATT GTCAGCCTGC TCAATGCCGA CGGTTTTATC AAAATCCTGA
251 CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCCGTTGGG AACGGTGTTG
301 GTTTCTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTCCGC
351 ATTAATGCGC TTATTGCTCA CAAAATCGCC ACGCAAATC ACTACTTTTA
401 TGGTTGTTTT TACAGGGATT TTATCTAATA CCGCTTCTGA ATTGGGCTAT
50  451 GTCGTCCTAA TCCCTTTGTC CGCCATCATC TTTCATTCCC TCGGCCGCCA
501 TCCGCTTGCC GGTCTGGCTG CGGCTTTCGC CGGCGTTTCG GGCGGTTATT

```

551	CGGCCAATCT	GTTCTTAGGC	ACAATCGATC	CGCTCTGGC	AGGCATCACC
601	CAACAGGCGG	CGCAAATCAT	CCATCCCAGC	TACGTCGTAG	GCCCTGAAGC
651	CAACTGGTTT	TTTATGGTAG	CCAGTACGTT	TGTGATTGCT	TTGATTGGTT
701	ATTTTGTTAC	TGAAAAATC	GTCGAACCGC	AATTGGGCCC	TTATCAATCA
751	GATTTGTAC	AAGAAGAAAA	AGACATTCCG	CATTCCAATG	AAATCAGCCG
801	TTTGAATAAT	AAAGGATTAA	TTTGGGCTGG	CGTGGTGTTT	GTTGCGCTTAT
851	CCGCCCTATT	GGCTTGGAGC	ATCGTCCCTG	CCGACGGTAT	TTTGCGT CAT
901	CTGAAACAG	GATTGGTTTC	CGGTTCCGCG	TTTTTAAAAAT	CGATTGTTGT
951	TTTTATTTTC	TTGTTGTTTG	CAGTGCCGGG	CATTGTTTAT	GGCCGGGTAA
1001	CCCGAAGTTT	CGCGGCGGAA	CAGGAAGTCG	TTAATCGCAT	GGCCGAATCG
1051	ATGAGTACTC	TGGGGCTTTA	TTTGGTCATC	ATCTTTTTTG	CCGCAAGATT
1101	TGTCGCATTT	TTTAATTGGA	CGAATATTGG	GCAATATATT	GCCGTTAAAG
1151	GGCGACGTT	CTTAAAAGAA	GTCGGCTTGG	GCGGACGCGT	GTTGTTTATC
1201	GGTTTTATTT	TAATTTGTGC	TTTTATCAAT	CTGATGATAG	GCTCCGCTTC
1251	CGCGCAATGG	CGGGTAACTG	CGCCGATTTT	CGTCCCATATG	CTGATTTTGG
1301	CCGGCTACGC	GCCCGAAGTC	ATTC AAGCCG	CTTACCGCAT	CGGTGATTCC
1351	GTTACCAATA	TTATTACGCC	GATGATGAGT	TATTTCCGGC	TGATTATGGC
1401	GACGGTGATC	AAATACAAAA	AAGATGCCGG	CGTGGGTACG	CTGATTCTTA
1451	TGATGTTGCC	GTATTTCCGCT	TTCTTCTTGA	TTGCGTGGAT	TGCCTTATTC
1501	TGCATTGGG	TATTTGTTTT	GGGCCTGCCC	GTCGGTCCCC	GCGCGCCCAC
1551	ATCTCATCCC	GCACCTTAA			

25

1	MSQTDTORQG	RFLRTVEWLG	NMLPHPVTLF	IIFIVLLLIA	SAVGAYFGLS
51	VPDPRPVGAK	GRADDGLIYI	VSLLNADGFI	KILTHTVKNF	TGFAPLGTVI
101	VSLGVGIAE	KSLGISALMR	LLLTSPRKL	TTFMVVFTGI	LSNTASELGY
151	VVLIPLSAII	FHSLGRHPLA	GLAAAFAGVS	GGYSANFLFG	TIDPLLAGIT
201	QQAQIIHPD	YVVGPEANWF	FMASTFVIA	LIGYFVTEKI	VEPQDLGYQS
251	DLSQEEKDIR	HSNEITPLEY	KGLIWAGVVF	VALSALLAWS	IVPADGILRH
301	PETGLVSGSP	FLKSIVVFIF	LLFALPGIVY	GRVTRSLRGE	QEVVNAMAES
351	MSTLGLYLV	IFFAAQFVAF	FNWNTNIGQY	AVKGATFLE	VGLGGSVLEI
401	GFLLICAFIN	LMIGSASAQW	AVTAPIFVPM	LMLAGYAPEV	IQAAAYRIGDS
451	VTNIITPMMS	YFGLIMATVI	KYKKDAGVGT	LISMMLPYSA	FFLIAWIALF
501	CWVFEVLGLP	VGPGAPTEYP	AP*		

30

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF12 shows 96.3% identity over a 320aa overlap with an ORF (ORF12a) from strain A of *N. meningitidis*:

40

```

                                10      20      30
orf12.pep                      AXXIIHPXXVVGPEANWFFMVASTFVIALI
                                |  |||  |||||  |||||  |||||  |||||
orf12a      AAFAAGVSGGYSANLFLGTIDPLLAGITQAAQIIHPDYVVGPEANWFFMVASTFVIALI
              180      190      200      210      220      230

```

45

```

                                40       50       60       70       80       90
orf12.pep    GYFVTEKIVEPQLGPYQSDLSQEKKDIRHSNEITPLEYKGLIWAGVVFFALSALLAWSIV
              |||||
orf12a       GYFVTEKIVEPQLGPYQSDLSQEKKDIRHSNEITPLEYKGLIWAGVVFFALSALLAWSIV
              240     250     260     270     280     290

```

50

```

      100      110      120      130      140      150
orf12.pep PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAXAESMS
          |||||
orf12a    PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAMAE SMS
          300      310      320      330      340      350

```

55

```

              160      170      180      190      200      210
orf12.pep  TLXLXLXXIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLM
            |||||
orf12a     TLGLYLVIIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLM
            |||||
              360      370      380      390      400      410

```

60

orf12.pep 220 230 240 250 260 270  
IGSASAQWAVTAPIFVPLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVXXY

-130-

```

orf12a      IGSASAQWAVTAPIFVPMMLLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKY
              420      430      440      450      460      470

5  orf12.pep      KKDAGVGTLLIXMMLPYSAFFLIAWIALFCIWVFLVGLPVGPGAPTFYPAPX
              280      290      300      310      320
orf12a      KKDAGVGTLLISMMLPYSAFFLIAWIALFCIWVFLVGLPVGPGAPTFYPAPX
              480      490      500      510      520

```

The complete length ORF12a nucleotide sequence <SEQ ID 137> is:

```

10      1  ATGAGTCAAA CCGATACGCA ACGGGACGGA CGATTTTAC GCACAGTCGA
      51  ATGGCTGGGC AATATGTTGC CGCACCCGGT TACGCTTTT ATTATTTTCA
     101  TTGTGTTATT GCTGATTGCC TCTGCCGCG GTGCGTATT CCGACTATCC
     151  GTCCCCGATC CGCGCCCTGT TGGTGCGAAA GGACGTGCCG ATGACGGTTT
     201  GATTCACGTT GTCAGCCTGC TCGATGCTGA CGGTTTGATC AAAATCCTGA
     15      251  CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCCGTGGG AACGGTGTG
     301  GTTCTTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTCCGC
     351  ATTAATGCGC TTATTGCTCA CAAAATCTCC ACGCAAATC ACTACTTTTA
     401  TGTTGTTTT TACAGGGATT TTATCTAATA CCGCTTCTGA ATTGGGCTAT
     451  GTCGTCCTAA TCCCTTTGTC CGCCATCATC TTTCATTCCC TCGGCCGCCA
     20      501  TCCGCTTGCC GGTCTGGCTG CGGCTTTCGC CGGCGTTTCG GCGGTTATT
     551  CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTGGC AGGCATCACC
     601  CAACAGGCGG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
     651  CAACTGGTTT TTTATGGTAG CCAGTACGTT TGTGATTGCT TTGATTGGTT
     701  ATTTTGTTAC TGAATAATC GTCGAACCGC AATTGGGCC TTATCAATCA
     25      751  GATTTGTAC AAGAAGAAA AGACATTCGA CATTCCAATG AAATCAGCC
     801  TTTGGAATAT AAAGGATTAA TTTGGGCTGG CGTGGTGTT GTTGCCTTAT
     851  CCGCCCTATT GGCTTGGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
     901  CCTGAAACAG GATTGGTTTC CGGTCGCCG TTTTAAAT CAATTGTTGT
     951  TTTATTTTC TTGTTGTTG CACTGCCGGG CATTGTTTAT GGCCGGTAA
     30      1001  CCCGAAGTTT GCGCGCGCAA CAGGAAGTCG TTAATGCGAT GGCCGAATCG
     1051  ATGAGTACTC TGGGGCTTTA TTTGGTCATC ATCTTTTTC CCGCACAGTT
     1101  TGTGCGATTT TTAATTGGA CGAATATTGG GCAATATATT GCCGTAAAG
     1151  GGGCGACGTT CTAAAAGAA GTCGGCTTGG GCGGCAGCGT GTTGTATATC
     1201  GGTTTTATT TAATTGTGC TTTTATCAAT CTGATGATAG GCTCCGCCTC
     35      1251  CGCGCAATGG GCGGTAAGT CCGCGATTT CGTCCCTATG CTGATGTGG
     1301  CCGGCTACGC GCCCGAAGTC ATTCAAGCCG CTTACCGCAT CGGTGATTCC
     1351  GTTACCAATA TTATTACGCC GATGATGAGT TATTCGGGC TGATTATGGC
     1401  GACGGTGATC AAATACAAA AAGATGCGGG CGTGGGTACG CTGATTCTA
     1451  TGATGTTGCC GTATTCCGCT TTCTTCTGA TTGCGTGGAT TGCCTTATC
     40      1501  TGCATTTGGG TATTTGTTTT GGGCCTGCCC GTCGGTCCC GCGCGCCAC
     1551  ATTCTATCCC GCACCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 138>:

```

45      1  MSQTDTRQDG RFLRTVEWLG NMLPHVPTLF IIFIVLLLIA SAAGAYFGLS
      51  VPDPRPVGAK GRADDGLIHV VSLLDADGLI KILHTVKNF TGFAPLGTVL
     101  VSLGVGIAE KSLISALMR LLLTKSPRKL TTFMVFTGI LSNTASELGY
     151  VVLPLSAII FHSLSRHPLA GLAAAFAGVS GGYSANFLG TIDPLLAGIT
     201  QQAQIIHPD YVVGPEANWF FMVASTFVIA LIGYFVTEKI VEPQLGPYQS
     251  DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
     301  PETGLVSGSP FLKSIVVFIF LLFALPGIVY GRVTRSLRGE QEVVNMAES
     50      351  MSTLGLYLV IFFAAQFVAF FNWTNIGQYI AVKGATFLKE VGLGGSVLEI
     401  GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGYAPEV IQAAYRIGDS
     451  VTNIITPMS YFGLIMATVI KYKKDAGVGT LISMMLPYSA FFLIAWIALF
     501  CIWVFLGLP VGPGAPTFYP AP*

```

55 ORF12a and ORF12-1 show 99.0% identity in 522 aa overlap:

```

60      orf12a.pep      MSQTDTRQDGRFLRTVEWLG NMLPHVPTLF IIFIVLLLIASAAGAYFGLSVDPDRPVGAK
              10      20      30      40      50      60
      orf12-1      MSQTDTRQDGRFLRTVEWLG NMLPHVPTLF IIFIVLLLIASAVGAYFGLSVDPDRPVGAK
              10      20      30      40      50      60

      orf12a.pep      GRADDGLIHVVSLLDADGLIKILHTVKNFTGFAPLGTVLVSLGVGIAEKSLISALMR
              70      80      90      100      110      120
      orf12-1      GRADDGLIYIVSLLDADGFIKILHTVKNFTGFAPLGTVLVSLGVGIAEKSLISALMR

```



-131-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf12a.pep	LLLT	KSPR	KLTT	FMVV	FTGIL	SNTASE
	orf12-1	LLLT	KSPR	KLTT	FMVV	FTGIL	SNTASE
		130	140	150	160	170	180
10	orf12a.pep	GGYS	ANLFL	GTID	PLLAG	ITQQA	QIIHP
	orf12-1	GGYS	ANLFL	GTID	PLLAG	ITQQA	QIIHP
		190	200	210	220	230	240
15	orf12a.pep	VEPQ	LGPY	QSDLS	QEEK	DIRHS	NEITP
	orf12-1	VEPQ	LGPY	QSDLS	QEEK	DIRHS	NEITP
		250	260	270	280	290	300
20	orf12a.pep	PETGL	VSGSP	FLKS	IVVF	IFLL	FALPG
	orf12-1	PETGL	VSGSP	FLKS	IVVF	IFLL	FALPG
		310	320	330	340	350	360
25	orf12a.pep	PETGL	VSGSP	FLKS	IVVF	IFLL	FALPG
	orf12-1	PETGL	VSGSP	FLKS	IVVF	IFLL	FALPG
		310	320	330	340	350	360
30	orf12a.pep	IFFAA	QFVA	FFNWT	NIGQY	IAVK	GATFL
	orf12-1	IFFAA	QFVA	FFNWT	NIGQY	IAVK	GATFL
		370	380	390	400	410	420
35	orf12a.pep	AVTAP	IFVPM	LMLAG	YAPE	VIQAA	YRIGD
	orf12-1	AVTAP	IFVPM	LMLAG	YAPE	VIQAA	YRIGD
		430	440	450	460	470	480
40	orf12a.pep	LISM	MLPY	SAFFL	IAWIA	LFCIW	VFVLG
	orf12-1	LISM	MLPY	SAFFL	IAWIA	LFCIW	VFVLG
		490	500	510	520		

45 Homology with a predicted ORF from *N.gonorrhoeae*

ORF12 shows 92.5% identity over a 320aa overlap with a predicted ORF (ORF12.ng) from *N. gonorrhoeae*:

	orf12.pep	AXXI	IHPX	XVVG	PEAN	WFFM	VAST	FVIALI	30
50	orf12ng	AAAF	AGVS	GGYS	ANLFL	GTID	PLLAG	ITQQA	QIIHP
	orf12.pep	GYFV	TEKIV	EPQL	GPYQ	SDLS	QEEK	DIRHS	NEITP
55	orf12ng	GYFV	TEKIV	EPQL	GPYQ	SDLS	QEEK	DIRHS	NEITP
	orf12.pep	PADG	ILRHP	PETGL	VSGSP	FLKS	IVVF	IFLL	FALPG
	orf12ng	PADG	ILRHP	PETGL	VAGSP	FLKS	IVVF	IFLL	FALPG
60	orf12.pep	TLXL	LXXI	FFAA	QFVA	FFNWT	NIGQY	IAVK	GATFL
	orf12ng	TLGL	YLVI	IFFA	QFVA	FFNWT	NIGQY	IAVK	GAVFL
65	orf12.pep	IGSAS	AQWAV	TAPI	FVPM	LMLAG	YAPE	VIQAA	YRIGD
	orf12ng	IGSAS	AQWAV	TAPI	FVPM	LMLAG	NAPQ	VIQAA	YRIGD

-132-

```

orf12.pep      KKDAGVGTLLXMMPLPYSAFFLIAWIALFCIWVFLGLPVGPGAPTFFYPAP 320
|||||:|||||:
orf12ng        KKDAGVGTLLSMMPLPYSAFFLIAWIALFCIWVFLGLPVGPGTPTFFYPVP 522

```

The complete length ORF12ng nucleotide sequence <SEQ ID 139> is:

```

5      1  ATGAGTCAAA CCGACGCGCG TCGTAGCGGA CGATTTTAC GCACAGTCGA
      51  ATGGCTGGGC AATATGTTGC CGCACCCGGT TACGCTTTT ATTATTTTCA
     101  TTGTGTTATT GCTGATTGcc tctgCCGTCG GTGCGTATT CGGACTATCC
     151  GTCCCCGATC CGCGTCCTGT TGGGGCGAAA GGACGTGCCG ATGACGGTTT
     201  GATTACAGTT GTCAGCCTGC TCGATGCCGA CGGTTTGATC AAAATCGCTGA
    10  251  CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCCGTTGGG AACGGTGTG
     301  GTTCTTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTCCGC
     351  ATTAATGCGC TTATTGCTCA CAAAATCCCC ACGCAAACCTC ACTACTTTTA
     401  TGGTTGTTTT TACAGGGATT TTATCCAATA CGGCTTCTGA ATTGGGCTAT
     451  GTCGTCCTAA TCCCTTTGTC CGCCGTCATC TTTCATTGCG TCGGCGCCCA
    15  501  TCCGCTTGCC GGTTTGGCTG CGGCTTTCGC CGGCGTTTCG GCGGTTATT
     551  CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTGGC AGGCATCACC
     601  CAACAGGCGG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
     651  CAACTGGTTT TTTATGGCAG CCAGTACGTT TGTGATTGCT TTGATTGGTT
     701  ATTTTGTTAC TGA AAAAATC GTCGAACCGC AATTGGGCC TTATCAATCA
    20  751  GATTTGTAC AAGAAGAAA AGACATTCGG CATTCCAATG AAATCACGCC
     801  TTTGGAATAT AAAGATTAA TTTGGGCAGG CGTGGTGTG GTTGCCATTAT
     851  CCGCCCTATT GGCTTGGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
     901  CCTGAAACAG GATTGGTTGC CGGTCGCGG TTTTAAAT CGATTGTGTG
     951  TTTATTTTC TTGTTGTTG CGTGCCGGG CATTGTTTAT GGCCGATAA
    25 1001  CCCGAAGTTT CGCGGCGGAA CGGGAAGTCG TTAATGCGAT GGCCGAATCG
     1051  ATGAGTACTT TGGGACTTTA TTTGGTCATC ATCTTTTGT CCGCACAGTT
     1101  TGTCGCATTT TTAATTGGA CGAATATTGG GCAATATATT GCCGTAAAG
     1151  GGGCGGTGTT CTAAAAGAA GTCGGCTTGG GCGGCAGTGT GTTGTATTATC
     1201  GGTTTTATT TAATTGTGC TTTTATCAAT CTGATGATAG GCTCCGCTC
    30 1251  CGCGCAATGG CGCGTAAGTC CGCGGATTT CGTCCCTATG CTGATGTTGG
     1301  CCGGCTACGC GCCCGAAGTC ATTCAAGCCG CTTACCGCAT CGGTGATTCC
     1351  GTTACCAATA TTATTACGCC GATGATGAGT TATTCGGGC TGATTATGGC
     1401  GACGGTAATC AAATACAAA AAGATGCGGG CGTAGGCACG CTGATTCTA
     1451  TGATGTTGCC GTATTCCGCT TCCTTCTTAA TTGCATGGAT CGCCTTATTC
    35 1501  TGCATTGGG TATTGTTTT GGGTCTGCCC GTCGGTCCCG GCACCCCAC
     1551  ATTCTATCCG GTGCCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 140>:

```

      1  MSQTDARRSG RFLRTVEWLG NMLPHPVTLF IIFIVLLLIA SAVGAYFGLS
     51  VPDPRPVGAK GRADDGLIHV VSLLDADGLI KILHTVKNF TGFAPLGTVL
    40 101  VSLLGVGIAE KSLGISALMR LLLTKSPRKL TTFMVFTGI LSNTASELGY
     151  VVLIPLSAVI FHSLSGRHPLA GLAAAFAGVS GGYSANFLG TIDPLLAGIT
     201  QQAAQIIHPD YVVGPEANWF FMAASTFVIA LIGYFVTEKI VEPQLGPYQS
     251  DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
     301  PETGLVAGSP FLKSIVVFIF LLFALPGIVY GRITRSLRGE REVVNAMAES
    45 351  MSTLGLYLVI IFFAAQFVAF FNWTNIGQYI AVKGAFLKK FRLGGSVLEI
     401  GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGNAPQV IQAAYRIGDS
     451  VTNIITPMMS YFGLIMATVI KYKKDAGVGT LISMMPLPYSA FFLIAWIALF
     501  CIWVFLGLP VPGTPTFFYP VP*

```

ORF12ng shows 97.1% identity in 522 aa overlap with ORF12-1:

```

50      10      20      30      40      50      60
orf12-1.pep  MSQTD TQRDGRFLRTVEWLG NMLPHPVTLF IIFIVLLLIASAVGAYFGLSV PDP RPVGAK
orf12ng      MSQTDARRSGRFLRTVEWLG NMLPHPVTLF IIFIVLLLIASAVGAYFGLSV PDP RPVGAK
      10      20      30      40      50      60
55      70      80      90      100     110     120
orf12-1.pep  GRADDGLIYIVSLLNADGFIKILHTVKNF TGFAPLGTVLV SLLGVGIAE KSLGISALMR
orf12ng      GRADDGLIHVVSLLDADGLIKILHTVKNF TGFAPLGTVLV SLLGVGIAE KSLGISALMR
      70      80      90      100     110     120
60      130     140     150     160     170     180
orf12-1.pep  LLLTKSPRKL TTFMVFTGILSNTASELGYV VLIPLSAIIFHSLSGRHPLAGLAAAFAGVS
orf12ng      LLLTKSPRKL TTFMVFTGILSNTASELGYV VLIPLSAIIFHSLSGRHPLAGLAAAFAGVS
      130     140     150     160     170     180
65

```

-133-

		130	140	150	160	170	180
		190	200	210	220	230	240
5	orf12-1.pep	GGYSANLFLGTIDPELLAGITQAAQIIHPDYVVGPEANWFFMVASTFVIALIGYFVTEKI					
	orf12ng	GGYSANLFLGTIDPELLAGITQAAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKI					
		190	200	210	220	230	240
10	orf12-1.pep	VEPQLGPIYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH					
	orf12ng	VEPQLGPIYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH					
		250	260	270	280	290	300
15	orf12-1.pep	PETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAMAESMSTLGLYLVI					
	orf12ng	PETGLVAGSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNAMAESMSTLGLYLVI					
		310	320	330	340	350	360
20	orf12-1.pep	IFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFILICAFINLMIGSASAQW					
	orf12ng	IFFAAQFVAFFNWTNIGQYIAVKGAVFLKEVGLGGSVLFIFILICAFINLMIGSASAQW					
		370	380	390	400	410	420
25	orf12-1.pep	AVTAPIFVPMMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT					
	orf12ng	AVTAPIFVPMMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT					
		430	440	450	460	470	480
30	orf12-1.pep	LISMMLPYSAFFLIAWIALFCIWWFVLGLPVGPGAPTFFYPAPX					
	orf12ng	LISMMLPYSAFFLIAWIALFCIWWFVLGLPVGPGTPTFFYPVPX					
		490	500	510	520		
35	orf12-1.pep						
	orf12ng						

In addition, ORF12ng shows significant homology with a hypothetical protein from *E.coli*:

40	sp P46133 YDAH_ECOLI HYPOTHETICAL 55.1 KD PROTEIN IN OGT-DBPA INTERGENIC REGION >gi 1787597 (AE000231) hypothetical protein in ogt 5'region [Escherichia coli] Length = 510 Score = 329 bits (835), Expect = 2e-89 Identities = 178/507 (35%), Positives = 281/507 (55%), Gaps = 15/507 (2%)
45	Query: 8 RSGRFLRTVEWLGNNMLPHPVTTXXXXXXXASAVGAYFGLSVDPDRPVGAKGRADDGL 67 +SG+ VE +GN +PHP +A+ +FG+S +P D Sbjct: 13 QSGKLYGWVERIGNKVPHFLLFIYLIIVLMVTTAILSAFGVSAKNP-----TDGTP 64
50	Query: 68 IHVVSLLDADGLIKILTHTVKNFTGFAPXXXXXXXXXXXXXIAEKSLISALMRLLLTKSP 127 + V +LL +GL L + +KNF+GFAP +AE+ GL+ ALM + + Sbjct: 65 VVVKNLLSVEGLHWFLPNVIKNFSGFAPLGAILALVLGAGLAERVGLLPALMVKMASHVN 124
55	Query: 128 RKLTTFMVFTGILSNTASELGYYVLIPLSAVIFHSLGRHPLAGLAAAFAGVSGGYSANL 187 + ++MV+F S+ +S+ V++ P+ A+IF ++GRHP+AGL AA AGV G++ANL Sbjct: 125 ARYASYMVLFIAFFSHISSDAALVIMPPMGALIFLAVGRHPVAGLLAAIAGVCGFTANL 184
60	Query: 188 FLGTIDPELLAGITQAAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKIVEPQLGP 247 + T D LL+GI+ +AA +P V NW+FMA+S V+ ++G +T+KI+EP+LG Sbjct: 185 LIVTTDVLLSGISTEAAAFNPQMHVSVIDNWFMASSVVVLTIVGGLITDKIIEPRLGQ 244
65	Query: 248 YQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRHPETGLVA 307 +Q + ++ + + S GL AGVV + A +A ++P +GILR P V Sbjct: 245 WQGNSEKLQTLTESQRF-----GLRIAGVVSLLFIAAIALMVIPQNGILRDPINHTVM 298
70	Query: 308 GSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNAMAESMSTLGLYLXXXXXXX 367 SPF+K IV I L F + + YG TR++R + ++ + M E M + ++ Sbjct: 299 PSPFIKIVPLIILFFVVSLEYGIATRTIRROADLPHLMIEPMKEMAGFIVMVFPPLAQF 358
	Query: 368 XXXXNWTNIGQYIAVKGAVFLKEVGLGGSVLFIFILICAFINLMIGSASAQWAVTAPIF 427 NW+N+G++IAV L+ GL G F+G L+ +F+ + I S SA W++ APIF

-134-

Sbjct: 359 VAMFNWSNMGKFIAVGLTDILESSGLSGIPAFVGLALLSSFLCMFIASGSAIWSILAPIF 418

Query: 428 VPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGTLSMMLP 487  
 VPM ML G+ P Q +RI DS + P+ + L + + +YK DA +GT S++LP

Sbjct: 419 VPMFMLLGFHPAFAQILFRIADSSVLPAPVSPFVPLFLGLFQRYKPDALGTYISLVLP 478

Query: 488 YSAFFLIAWIALFCIWVFLVGLPVGPG 514  
 Y FL+ W+ + W +++GLP+GPG

Sbjct: 479 YPLIFLVVWLLMLLAW-YLVGLPIGPG 504

Based on this analysis, including the presence of several putative transmembrane domains and the predicted actinin-type actin-binding domain signature (shown in bold) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 17

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 141>:

```

1  ..ACAGCCGGCG CAGCAGGTTn CnCGGTCTTC GTTTTCGTAA CGGACAGTCA
51  GGTGGAGGTG TTCGGGAACA TCCAGACCGC AGTGGAACA GGTTTTTTTC
101 ATGGCATTTC GGTTCGTCT GTGTTTGGTG CGGCGGCACA AGACTCGGCA
151 ATgGCTTCGC GCAGTGCCTC TATACCGGTA TTTTCAGCAA CGGAAATGCG
201 GACGCGcGgCA ATTTTCCCG CAGCGTCGCG CCATATGCCC GTGTTTTgTT
251 CTTCAGACGG CAGCAGGTCG GTTTGTGTGT ACACCTTgAT GCACGGAaTA
301 TCGCCGGCAT GGATTTCTTG CAGTACGTTT TCCACGTCTT CAATCTGCTG
351 TCCGCTGTTC GGAGCGGCGG CATCGACGAC GTGCAGCAGC ACATCgGcTT
25  401 gCGCGGTTTC TTCCAGCGTG GCgGAAAAGG CGGAAATCAG TTTgTGCGGC
451 agATyGCTnA CGAATCCGAC GGTATCGGTC AGGATAATGC TGCATTGCGG
501 ACT..

```

This corresponds to the amino acid sequence <SEQ ID 142; ORF14>:

```

1  ..TAGAAGXXVF VFVTDQVEV FGNIQTAVET GFFHGISVSS VFGAAAQDSA
51  MASRSASIPV FSATEMRTAA IFPAASRHMP VFCSSDGSRS VLLYTLMHGI
101 SPAWISCSTF STSSICPLF GAAASTTCSS TSACAVSSSV AEKAEISLCG
151 RXLTNPVSV RIMLHSG..

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF14 shows 94.0% identity over a 167aa overlap with an ORF (ORF14a) from strain A of *N.meningitidis*:

```

                                10      20      30
orf14.pep                      TAGAAGXXVFVFVTDQVEVFGNIQTAVET
                                |::|||  |||||::|::|::|  |||
40  orf14a      GRQLGFLRVGGALFVITAQARVNNALCDCLTTGAAGFAVFVFDGQMVFQVPAVET
                   150      160      170      180      190      200

                                40      50      60      70      80      90
orf14.pep      GFFHGISVSSVFGAAAQDSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS
45  orf14a      GFFHGISVSSVFGAAAQYSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS
                   210      220      230      240      250      260

                                100      110      120      130      140      150
50  orf14.pep      VLLYTLMHGISPWISCSTFSTSSICPLFGAAASTTCSSTSACAVSSSVAEKAEISLCG
                   |||||  |||||  |||||  |||||  |||||  |||||
orf14a      VLLYTLMHGISPWISCSTFSTSSICPLFGAAASTTCSSTSACAVSSSVAEKAEISLCG
                   270      280      290      300      310      320

```

```

                                160
orfl4.pep    RXLTNPTVSVRIMLHSG
              | | | | | | | | | |
orfl4a       RSLTNPTVSVRIMLHSGLMYSRRRAVVSSVAKSWSFAYMPDLVSRLNRLDLPVLVX
5            330      340      350      360      370      380

```

The complete length ORF14a nucleotide sequence <SEQ ID 143> is:

	1	ATGGAGGATT	TGCAGGAAAT	CGGGTTCGAT	GTCGCCGCCG	TAAAGGTAGG
	51	TCGGCAGCGC	GAACATCATC	GTCTGCATCA	TCCCAGCCCC	GGCAACGGCG
10	101	AGGCGGACGA	TGTATTGTTT	GCGTTCCTTT	TGGTTGCGCG	CTTCGATTTT
	151	TTGCGCGTCA	TAGGGTGCGG	CGGTGTAGCC	TATCTGCCTG	ATTTTCAACA
	201	GAATGTCGGA	AAGGCGGATT	TTGCCGTCGT	CCGACAGCAC	GCGCGACCGG
	251	TGCGTGCTGT	AATTGAGGTC	GATGCGGACG	ATGCCGTCTG	TACGCAAAAG
	301	CTGCTGTTTC	ATCAGCCAGA	CGCAGGCGGC	GCAGGTGATG	CCGCCGAGCA
15	351	TTAAAACCGC	CTCGCGCGTG	CCGCCGTGGG	TTTCCACAAA	GTCGGACTGG
	401	ACTTCGGGGA	GGTCGTACAG	GCGGATTTGG	TCGAGGATTT	CTTGGGGCGG
	451	CAGCTCGGTT	TTTTGCGCGT	CGGCGGTGCG	TTGTTTGTA	TAAC TGCCCA
	501	AGCCCGCGTC	AATAATGCTT	TGTGCGACTG	CCTGACAACC	GGCGCAGCAG
	551	GTTTCGCGGT	CTTCGTTTTC	GTAACGGACG	GTCAGATGCA	GGTTTTCGGG
20	601	AACGTCAGC	CCGCACTGGA	AACAGGTTTT	TTTCACTGGA	TTTCGGTTTT
	651	GTCTGTGTTT	GGTGCGGCGG	CACAATACTC	GGCAATGGCT	TTCGCGAGTG
	701	CGTCTATACC	GGTATTTTCA	GCAACGGAAA	TGCGGACGGC	GGCAATTTTT
	751	CCCGCAGCGT	CGCGCCATAT	GCCCGTGTTT	TGTTCTTCAG	ACGCGACGAG
	801	GTCGGTTTTG	TTGTACACCT	TGATGCACGG	AATATCGCCG	GCATGGATTT
25	851	CTTGCACTAG	GTTTCCACCG	TCTTCAATCT	GCTGTCGCTG	GTTCCGAGCG
	901	CGGGCATCGA	CGACGTGCAG	CAGCACATCG	GCTTGCGCGG	TTTCTTCCAG
	951	CGTGGCGGAA	AAGGCGGAAA	TCAGTTTGTC	CGGCAGATCG	CTGACGAATC
	1001	CGACGGTATC	GGTCAGGATA	ATGCTGCATT	CGGGACTGAT	GTACAGCCGC
	1051	CGCGCCGTCG	TGTCGAGTGT	GGCGAAAAGC	TGGCTTTTTC	CATATATGCC
	1101	CGACTTTGGTC	AGCCGGTTGA	ACAGACTTGA	TTTGCCCGACA	TTTGATATAG

30 This encodes a protein having amino acid sequence <SEQ ID 144>:

1	MEDLQEIGFD	VAAVKVGRQR	EHHRLHHPQP	GNGEADDVLF	AFFLVGGFDF
51	LRVIGCGGVA	YLPDFQQNVG	KADFAVVPDD	AAAVRAVIEV	DADDAVCTQK
101	LLFDQPDAGG	AGDAAEH*NR	LARAAGVFHK	VGLDFGQVVQ	ADLVEDFLGR
151	QLGFLRVGGA	LFVITAQARV	NNALCDCLTT	GAGFAVVFV	VTDGQMVFVG
201	NVQPAVETGF	FHGISVSSVF	GAAQYSAMA	SRSASIPVFS	ATEMRTAAIF
251	PAASRHMPVF	CSSDGSRSVL	LYTLMHGISP	AWISCSSTFS	SSICCPPLFGA
301	AASTTCSSTS	ACAVSSSVAE	KAEISLCGRS	LTNPTVSVRI	MLHSGLMYSR
351	RAVVSSVAKS	WSFAYMPDLV	SRNLRLDLPT	LV*	

It should be noted that this sequence includes a stop codon at position 118.

40 Homology with a predicted ORF from *N.gonorrhoeae*

ORF14 shows 89.8% identity over a 167aa overlap with a predicted ORF (ORF14.ng) from *N. gonorrhoeae*:

	orf14.pep	TAGAAGXXVFVFTDSQVEVFGNIQTAVET	30
45	orf14ng	GRQGFFRVGGASFVITAQAGIDDALCDCLTADAAGFAVFADFADGQMVFQVGNVQPAVET	208
	orf14.pep	GFFHGISVSSVFGAAQDSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS	90
50	orf14ng	GFFHGISVSSVFGAAQYSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS	268
	orf14.pep	VLLYTLMHGISPWWISCSTFSTSSICCPLEFGAAASTTCSSTSACAVSSSSVAEKAELSLCG	150
55	orf14ng	VLLYTLMHGISPWWISCSTFSTSSICCPLEFGAAASTTCSSTSACTVSSKVAEKAELSLCG	328
	orf14.pep	RXLNPTVSVRIMLHSG	167
	orf14ng	RSLNPTVSVRIMLHAGLMYSRRVVSRAVKSWSFAYMPDLVSRNLRLDLPTLV	382

The complete length ORF14ng nucleotide sequence <SEQ ID 145> is predicted to encode a protein having amino acid sequence <SEQ ID 146>:

```

1 MEDLQEIGFD VAAVKVGRQR EHHRLHHTQS GNGKADDVLF AFFLVGGFDF
51 LRVIGCGGVA CLPDFQQNVG EADFAVVPDD AAARAVIEV DADDAVCAQK
101 LLFDQPDAGG AGNAAEHQHC FVRAIMGFHK VGLDFGQVVQ ADLVEDFLGR
151 QGFFFRVGGA SFVITAQAGI DDALCDCLTA DAAGFAVFVF VADGQMVFVG
5 NVQPAVETGF FHGISVSSVF GAAQYSAMA SRSASIPVFS ATEMRTAAIF
201 PAASRHMPVF CSSDGSRSVL LYTLMHGISW AWISCSTFST SSICCPLFRA
251 AASTTCSSTS ACTVSSKVAE KAEISLCGRS LTNPVTSVRI MLHAGLMYSR
301 RAVVSRVAKS WSFAYMPDLV SRLNRLDLPT LV*
351

```

Based on the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 18

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 147>:

```

15 1 ..GGCCATTACT CCGACCGCAC TTGGAAGCCG CGTTTGGNCG GCCGCCGTCT
51 GCCGTATCTG CTTTATGGCA CGCTGATTGC GGTATTGTG ATGATTTTGA
101 TGCCGAATC GGGCAGCTTC GGTTCGGCT ATGCGTCGCT GGCGGCTTTG
151 TCGTTCGGCG CGCTGATGAT TGCCTGTGTA GACGTGTCGT CAAATATGGC
201 GATGCAGCCG TTAAAGATGA TGGTCGGCGA CATGGTCAAC GAGGAGCAGA
251 AAA.NTACGC CTACGGGATT CAAAGTTTCT TAGCAAATAC GGGCGCGGTC
20 GTGGCGGCGA TTCTGCCGTT TGTGTTGCG TATATCGGTT TGGCGAACAC
351 CGCCGANAAA GGCGTTGTGC CGCAGACCGT GGTCGTGGCG TTTTATGTGG
401 GTGCGGCGTT GCTGGTGATT ACCAGCGCGT TCACGATTTT CAAAGTGAAG
451 GAATACGANC CGGAAACCTA CGCCCGTTAC CACGGCATCG ATGTCGCCCG
501 GAATCAGGAA AAAGCCAAC GTATCGCACT CTTAAAA.CC GCGC..

```

25 This corresponds to the amino acid sequence <SEQ ID 148; ORF16>:

```

1 ..GHYSDRTWKP RLXGRRLPYL LYGTIAVIV MILMPNSGSF GFGYASLAAL
51 SFGALMIALL DVSSNMAMQP FKMMVGD MVN EEQKXYAYGI QSFLANTGAV
101 VAAILPFVFA YIGLANTAXK GVVPQTVVVA FYVGAALLVI TSAFTIFKVK
151 EYXPETYARY HGIDVAANQE KANWIALLLX A..

```

30 Further work revealed the complete nucleotide sequence <SEQ ID 149>:

```

35 1 ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC
51 AAAAAGCACG ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG
101 CCTTACCCT GCAAAGCTCG CAAATGAGCC GCATTTTCA AACGCTAGGC
151 GCAGACCCGC ACAATTTGGG CTGGTTTTTC ATCCTGCCGC CGCTGGCGGG
201 GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCCGACCGC ACTTGAAGC
251 CGCGTTTGGG CGGCCGCCGT CTGCCGTATC TGCTTTATGG CACGCTGATT
301 GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGGCAGCT TCGGTTTCGG
351 CTATGCGTCG CTGGCGGCTT TGTGTTTCGG CGCGCTGATG ATTGCGCTGT
401 TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAGAT GATGGTCGGC
40 GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT
501 CTTAGCAAAT ACGGGCGCGG TCGTGGCGGC GATTCTGCCG TTTGTGTTTG
551 CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCGAGACC
601 GTGGTCGTGG CGTTTTATGT GGGTGCGGCG TTGCTGTTGA TTACCAGCGC
651 GTTACGATT TTCAAAGTGA AGGAATACGA TCCGGAACC TACGCCGTT
45 ACCACGGCAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA
751 CTCTGAAAA CCGCGCCTAA GCGTTTTGG ACGGTTACTT TGGTGCAATT
801 CTTCTGCTGG TTCGCCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA
851 TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTCCGT AGGTATCAG
901 GAGGCGGGTA ACTGGTACGG CGTTTTGGCG GCGGTGCAGT CGGTTGCGGC
50 GGTGATTTGT TCGTTTGTAT TGGCGAAAGT GCCGAATAAA TACCATAAGG
1001 CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT
1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGCTTATA CCTTAATCGG
1101 CATCGCTTGG GCGGGCATTA TCACTTATCC GCTGACGATT GTGACCAACG
1151 CCTGTGCGGG CAAGCATATG GGCATTACT TGGGCTTGTT TAACGGCTCT
55 ATCTGTATGC CTCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC
1251 TATGCTGGGC GGCTGTCAGG CCACTATGTT CTTGGTAGGG GGCGTCGTCC
1301 TGCTGCTGGG CGGTTTTTCC GTGTTCTGA TTAAGAAAC ACACGGCGGG
1351 GTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 150; ORF16-1>:

```

      1  MSEYTPQTAK  QGLPALAKST  IWMLSFGLG  VQTAFTLQSS  QMSRIFQTLG
     51  ADPHNLGWFF  ILPPLAGMLV  QPIVGHYSR  TWKPRLGRR  LPYLLYGTLI
    101  AVIVMILMPN  SGSFGFGYAS  LAALSFGALM  IALLDVSSNM  AMQPFKMMVG
    151  DMVNEEQKGY  AYGIQSFLAN  TGAVVAAILP  FVFAYIGLAN  TAEKGVVPQT
    201  VVVAFYVGAA  LLVITSFTI  FKVKEYDPET  YARYHGIDVA  ANQEKANWIE
    251  LLKTAPKAFW  TVTLVQFFCW  FAFQYMWYTS  AGAIAENVWH  TTDASSVGYY
    301  EAGNWDYVLA  AVQSVAVIC  SFVLAKVPNK  YHKAGYFGCL  ALGALGFFSV
    351  FFIGNQYALV  LSYTLIGIAW  AGIITYPLTI  VTNALSGKHM  GTYLGFLNGS
    401  ICMPQIVASL  LSFVLFPMLG  GLQATMFLVG  GVVLLLGAFS  VFLIKETHGG
    451  V*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF16 shows 96.7% identity over a 181aa overlap with an ORF (ORF16a) from strain A of *N.*

15 *meningitidis*:

```

                                10      20      30
orfl6.pep                      GHYSDRTWKPRXLGRRLPYLLYGTLIIVIV
                                |||
20  orfl16a  IFQTLGADPHSLGWFFILPPLAGMLVQPIVGHYSDRTWKPRLGRRLPYLLYGTLIIVIV
                        50      60      70      80      90     100

                                40      50      60      70      80      90
orfl6.pep  MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDVNEEQKXYAYGI
25  orfl16a  MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDVNEEQKGYAYGI
                        110     120     130     140     150     160

                                100     110     120     130     140     150
30  orfl6.pep  QSFLANTGAVVAAILPFVFAVYIGLANTAXKGVVPQTVVVAFYVGAAALLVITSFTIFKVK
orfl16a  QSFLANTGAVVAAILPFVFAVYIGLANTAEEKGVVPQTVVVAFYVGAAALLVITSFTIFKVK
                        170     180     190     200     210     220

                                160     170     180
35  orfl6.pep  EYXPETYARYHGIDVAANQEKANWIALLKXA
orfl16a  EYNPETYARYHGIDVAANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWYTSAGAI
                        230     240     250     260     270     280

40  orfl16a  AENVWHHTDASSVGYYEAGNWDYVLAQVSVAAVICSFVLAKVPNKYHKAGYFGCLALGA
                        290     300     310     320     330     340

```

The complete length ORF16a nucleotide sequence <SEQ ID 151> is:

```

      1  ATGTCGGAAT  ATACGCCTCA  AACAGCAAAA  CAAGGTTTGC  CCGCGCTGGC
     51  AAAAAGCACG  ATTTGGATGC  TCAGTTTCGG  CTTTCTCGGC  GTTCAGACGG
    101  CCTTTACCCT  GCAAAGCTCG  CAGATGAGCC  GCATCTTCCA  GACGCTCGGT
    151  GCCGATCCGC  ACAGCCTCGG  CTGGTTCTTT  ATCCTGCCGC  CGCTGGCGGG
    201  GATGCTGGTG  CAGCCGATTG  TCGGCCATTA  CTCCGACCGC  ACTTGGAAGC
    251  CGCGTTTGGG  CGGCCGCGGT  CTGCCGTATC  TGCTTTATGG  CACGCTGATT
    301  CCGGTTATTG  TGATGATTTT  GATGCCGAAC  TCGGGCAGCT  TCGGTTTCGG
    351  GCTATGCTCG  CTGGCGGCTT  TGTGTTTCGG  CGCGCTGATG  ATTGCGCTGT
    401  TAGACGTGTC  GTCAATATG  GCGATGCAGC  CGTTTAAGAT  GATGGTCGGC
    451  GACATGTGCA  ACGAGGAGCA  GAAAGGCTAC  GCCTACGGGA  TTCAAAGTTT
    501  CTTAGCGAAT  ACGGGCGCGG  TCGTGGCGGC  GATTCTGCCG  TTTGTGTTTG
    551  CGTATATCGG  TTTGGCGAAC  ACCGCCGAGA  AAGCGGTTGT  GCCGCAGACC
    601  GTGGTCGTGG  CGTTTTATGT  GGGTGCGGCG  TTGCTGGTGA  TTACCAGCGC
    651  GTTCACGATT  TTCAAAGTGA  AGGAATACAA  TCCGGAACCC  TACGCCCATT
    701  ACCACGGCAT  CGATGTCGCC  GCGAATCAGG  AAAAAGCCAA  CTGGATCGAA
    751  CTCTTGAAAA  CCGCGCCTAA  GCGGTTTGG  ACGGTTACTT  TGGTGCAATT
    801  CTTCTGCTGG  TTCGCCTTCC  AATATATGTG  GACTTACTCG  GCAGGCGCGA
    851  TTGCGGAAAA  CGTCTGGCAC  ACCACCGATG  CGTCTTCCGT  AGGTTATCAG
    901  GAGGCGGGTA  ACTGGTACGG  CGTTTGGCG  GCGGTGCAGT  CGGTTGCGGC
    951  GGTGATTTGT  TCGTTTGTAT  TGGCGAAAGT  GCCGAATAAA  TACCATAAGG

```

-138-

1001 CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT  
 1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTTAATCGG  
 1101 CATCGCTTGG GCGGGCATA TCACCTATCC GCTGACGATT GTGACCAACG  
 1151 CTTGTGCGGG CAAGCATATG GGCCTTACT TGGGCCTGTT TAACGGCTCT  
 5 1201 ATCTGTATGC CGCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC  
 1251 TATGCTGGGC GGCTTGCAGG CCACTATGTT CTTGGTAGGG GGCGTCGTCC  
 1301 TGCTGCTGGG CGCGTTTTCC GTGTTCTCTGA TTAAAGAAAC ACACGGCGGG  
 1351 GTTTGA

This encodes a protein having amino acid sequence <SEQ ID 152>:

10 1 MSEYTPQTAK QGLPALAKST IWMLSFGFLG VQTAFTLQSS QMSRIFQTLG  
 51 ADPHSLGWFF ILPPLAGMLV QPIVGHYSR TWKPRLGRR LPYLLYGTLI  
 101 AVIVMILMPN SGSFGFGYAS LAALSFGALM IALLDVSSNM AMQPFKMMVG  
 151 DMVNEEQGY AYGIQSFLAN TGAVVAAILP FVFAYIGLAN TAEKGVVPQT  
 201 VVVAFYVGAA LLVITSFTI FKVKEYNPET YARYHGIDVA ANQEKANWIE  
 15 LLKTAPKAFW TVTLVQFFCW FAFQYMWYTS AGAIAENVWH TTDASSVGQY  
 301 EAGNWWYGLA AVQSVAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFFSV  
 351 FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGFLNGS  
 401 ICMPQIVASL LSFVLFPMGL GLQATMFLVG GVVLLGAFS VFLIKETHGG  
 451 V\*

20 ORF16a and ORF16-1 show 99.6% identity in 451 aa overlap:

		10	20	30	40	50	60
orf16a.pep		MSEYTPQTAKQGLPALAKSTIWMLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHSLGWFF					
orf16-1		MSEYTPQTAKQGLPALAKSTIWMLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFF					
25		10	20	30	40	50	60
orf16a.pep		ILPPLAGMLVQPIVGHYSRDTWKPRLGRRLPYLLYGTLIAVIVMILMPNSGSFGFGYAS					
orf16-1		ILPPLAGMLVQPIVGHYSRDTWKPRLGRRLPYLLYGTLIAVIVMILMPNSGSFGFGYAS					
30		70	80	90	100	110	120
orf16a.pep		LAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQGYAYGIQSFLANTGAVVAAILP					
orf16-1		LAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQGYAYGIQSFLANTGAVVAAILP					
35		130	140	150	160	170	180
orf16a.pep		FVFAYIGLANTAEGKGVVPQTVVVAFYVGAAALLVITSFTIFKVKEYNPETYARYHGIDVA					
orf16-1		FVFAYIGLANTAEGKGVVPQTVVVAFYVGAAALLVITSFTIFKVKEYNPETYARYHGIDVA					
40		190	200	210	220	230	240
orf16a.pep		ANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWYTSAGAIAENVWHTTDASSVGQY					
orf16-1		ANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWYTSAGAIAENVWHTTDASSVGQY					
45		250	260	270	280	290	300
orf16a.pep		EAGNWWYGLAAVQSVAVICSFVLAKVPNKYHKAGYFGCLALGALGFFSVFFIGNQYALV					
orf16-1		EAGNWWYGLAAVQSVAVICSFVLAKVPNKYHKAGYFGCLALGALGFFSVFFIGNQYALV					
50		310	320	330	340	350	360
orf16a.pep		LSYTLIGIAWAGIITYPLTIVTNALSGKHMGTYLGLFNGSICMPQIVASLLSFVLFPMGL					
orf16-1		LSYTLIGIAWAGIITYPLTIVTNALSGKHMGTYLGLFNGSICMPQIVASLLSFVLFPMGL					
55		370	380	390	400	410	420
orf16a.pep		GLQATMFLVGGVVLLGAFSVFLIKETHGGVX					
orf16-1		GLQATMFLVGGVVLLGAFSVFLIKETHGGVX					
60		370	380	390	400	410	420
orf16a.pep		GLQATMFLVGGVVLLGAFSVFLIKETHGGVX					
orf16-1		GLQATMFLVGGVVLLGAFSVFLIKETHGGVX					
65		430	440	450			
orf16a.pep		GLQATMFLVGGVVLLGAFSVFLIKETHGGVX					
orf16-1		GLQATMFLVGGVVLLGAFSVFLIKETHGGVX					



Homology with a predicted ORF from *N.gonorrhoeae*

ORF16 shows 93.9% identity over a 181aa overlap with a predicted ORF (ORF16.ng) from *N. gonorrhoeae*:

5	orf16.pep	GHYSDRTWKPRXLGRRLLPYLLYGTIAIV	30
	orf16ng	HFSNARRRPAQFGLVFHPAAAGGDAGSADSGYSDRTWKPRLLGRRLLPYLLYGTIAIV	131
10	orf16.pep	MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQXYAYGI	90
	orf16ng	MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKSYAYGI	191
	orf16.pep	QSFLANTGAVVAAILPFVFAYIGLANTAXKGVVPQTVVVAFYVGAALLVITSFTIFKVK	150
15	orf16ng	QSFLANTDAVVAAILPFVFAYIGLANTAEGVVPQTVVVAFYVGAALLIITSFTISKVK	251
	orf16.pep	EYXPETYARYHGIDVAANQEKANWIALLKXA	181
	orf16ng	EYDPETYARYHGIDVAANQEKANWFELLKTAPKVFWTVTPVQFFCWFAPRYMWTYSAGAI	311

20 The complete length ORF16ng nucleotide sequence <SEQ ID 153> is:

	1	ATGATAGGGG	ATCGCCGCGC	CGGCAACCAT	TTCGGATTTT	CCAAAGCAAA
	51	TACTTTTCAA	ATCAAAAAAA	AGGATTACT	TTATGTCGGA	ATATACGCCT
	101	CAACAGCAA	AACAAGGTTT	GCCCCGCGCG	GCAAAAAGCA	CGATTTGGAT
	151	GTTGAGCTTC	GGCTATCTCG	GCGTTCAGAC	GGCCTTTACC	CTGCAAAGCT
25	201	CGCAGATGAG	CCGCATTTTT	CAAACGCTAG	GCGCAGACCC	GCACAATTTG
	251	GGCTGGTTTT	TCATCCTGCC	GCCGCTGGCG	GGGATGCTGG	TTCAGCCGAT
	301	AGTGCTACT	ACTCAGACCG	CACTTGGAAG	CCGCGCTTGG	GCGGCCGCGG
	351	CCTGCCGTAT	CTGCTTTACG	GCACGCTGAT	TGCGGTCATC	GTGATGATTT
	401	TGATGCCGAA	CTCGGGCAGC	TTCGGTTTCG	GCTATGCGTC	GCTGGCGGCC
30	451	TTGTGCTTCG	GCGCGCTGAT	GATTGCGCTG	TTGGACGTGT	CGTCGAATAT
	501	GGCGATGCAG	CCGTTTAAGA	TGATGGTCGG	CGATATGGTC	AACGAGGAGC
	551	AGAAAAGCTA	CGCCTACGGG	ATTCAAAGTT	TCTTAGCGAA	TACGGACGCG
	601	GTTGTGGCAG	CGATTCTGCC	GTTTGTGTTC	GCGTATATCG	GTTTGGCGAA
	651	CACTGCCGAG	AAAGGCGTTG	TGCCACAAAC	CGTGGTCGTA	GCATTCTATG
35	701	TGGGTGCGGC	GTTACTGATT	ATTACCAAGT	CGTTCACAAAT	CTCCAAATAT
	751	AAAGAATACG	ACCCGGAAC	CTACGCCCGT	TACCACGGCA	TCGATGTCGC
	801	CGCGAATCAG	GAAAAGCCA	ACTGGTTCGA	ACTCTTAAAA	ACCGCGCCTA
	851	AAGTGTTTTG	GACGGTTACT	CCGGTACAGT	TTTTCTGCTG	GTTTCGCCTC
	901	CGGTATATGT	GGACTTACTC	GGCAGGCGCG	ATTGCAGAAA	ACGCTGGGCA
40	951	CACTACCGAT	GCGTCTTCCG	TAGGCCATCA	GGAGGCGGGC	AACCGGTACG
	1001	GCGTTTTGGC	GGCGGTGTAG			

This encodes a protein having amino acid sequence <SEQ ID 154>:

	1	MIGDRRAGNH	FGFSKANTFQ	IKKKDLLYVG	IYASNSKTRF	ARAGKKHDL
	51	VELRLSRRSD	GLYPAKLADE	PHFSNARRRP	AQFGLVFHFA	AAGGDAGSAD
45	101	SGYSDRTWK	PRLGRRLLPY	LLYGTIAVI	VMILMPNSGS	FGFGYASLAA
	151	LSFGALMIAL	LDVSSNMAMQ	PFKMMVGDMV	NEEQKSYAYG	IQSFLANTDA
	201	VVAAILPFVF	AYIGLANTAE	KGVVPQTVV	AFYVGAALLI	ITSFTISKV
	251	KEYDPETYAR	YHGIDVAANQ	EKANWFELLK	TAPKVFWTVT	PVQFFCWFAP
	301	RYMWTYSAGA	IAENVHHTD	ASSVGHQEAG	NRYGVLAAV*	

50 ORF16ng and ORF16-1 show 89.3% identity in 261 aa overlap:

		30	40	50	60	70	80
	orf16-1.pep	MLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFFILPPLAGMLVQPI-VGHYSDRT					
	orf16ng	DVELRLSRRSDGLYPAKLADEPHFSNARRRPAQFGLVF-HPAAGGDAGSADSGYSDRT					
55		50	60	70	80	90	100
	orf16-1.pep	WKPRLLGRRLLPYLLYGTIAIVVMILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMA					
	orf16ng	WKPRLLGRRLLPYLLYGTIAIVVMILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMA					
60		110	120	130	140	150	160

-140-

		150	160	170	180	190	200
5	orf16-1.pep	MQPFKMMVGDMDVNEEQKGYAYGIQSFLANTGAVVAAILPFVFAYIGLANTAEGVVPQTV					
	orf16ng	MQPFKMMVGDMDVNEEQKSYAYGIQSFLANTDAVVAAILPFVFAYIGLANTAEGVVPQTV					
		170	180	190	200	210	220
10	orf16-1.pep	VVAFYVGAALLVITSAFTIFKVKEYDPETYARYHGIDVAANQEKANWIELLKTA PKAFWT					
	orf16ng	VVAFYVGAALLIITSAFTISKVKEYDPETYARYHGIDVAANQEKANWFELLKTA PKVFWT					
		230	240	250	260	270	280
15	orf16-1.pep	VTLVQFFCWFQYMWYTYSAGAIAENVWHTTDASSVGYQEAGNWWYGVLA AVQSVAVICS					
	orf16ng	VTPVQFFCWFQYMWYTYSAGAIAENVWHTTDASSVGHQEAGNRYGVLA AVX					
		290	300	310	320	330	340

- 20 Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 19

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 155>:

25	1	ATGTTGTTCC	GTAAAACGAC	CGCCGCCGTT	TTGGCGCATA	CCTTGATGCT
	51	GAACGGCTGT	ACGTTGATGT	TGTGGGGAAT	GAACAACCCG	GTCAGCGAAA
	101	CAATCACCCG	NAAACACGTT	GNCAAAGACC	AAATCCGNGN	CTTCGGTGTG
	151	GTTGCCGAAG	ACAAATGCCCA	ATTGGAAAAG	GGCAGCCTGG	TGATGATGGG
30	201	CGGAAAATAC	TGGTTCGTCG	TCAATCCCGA	AGATTCGGCG	AA.NTGACGG
	251	GNATTTTGAN	GGCAGGGCTG	GACAAACCCT	TCCAAATAGT	TNAGGATACC
	301	CCGAGCTATG	C.TGCCACCA	AGCCCTGCCG	GTCAAACCTG	GATCGNCTGG
	351	CAGCCAGAAT...				

This corresponds to the amino acid sequence <SEQ ID 156; ORF28>:

35	1	MLFRKTTAAV	LAHTLMLNGC	TLMLWGMNPN	VSETITRKHV	XKDQIRXFGV
	51	VAEDNAQLEK	GSLVMMGGKY	WFFVNPEDSA	XXTGILXAGL	DKPFQIVXDT
	101	PSYXCHQALP	VKLGSXGSQN...			

Further work revealed the complete nucleotide sequence <SEQ ID 157>:

40	1	ATGTTGTTCC	GTAAAACGAC	CGCCGCCGTT	TTGGCGGCAA	CCTTGATGCT
	51	GAACGGCTGT	ACGTTGATGT	TGTGGGGAAT	GAACAACCCG	GTCAGCGAAA
	101	CAATCACCCG	CAAACACGTT	GACAAAGACC	AAATCCGCGC	CTTCGGTGTG
	151	GTTGCCGAAG	ACAAATGCCCA	ATTGGAAAAG	GGCAGCCTGG	TGATGATGGG
	201	CGGAAAATAC	TGGTTCGTCG	TCAATCCCGA	AGATTCGGCG	AAGCTGACGG
	251	GCATTTTGAA	GGCAGGGCTG	GACAAACCCT	TCCAAATAGT	TGAGGATACC
45	301	CCGAGCTATG	CTCGCCACCA	AGCCCTGCCG	GTCAAACCTG	AATCGCCTGG
	351	CAGCCAGAAT	TTCAGTACCG	AAGGCCTTTG	CCTGCGCTAC	GATACCGACA
	401	AGCCTGCCGA	CATCGCCAAG	CTGAAACAGC	TCGGGTTTGA	AGCGGTCAAA
	451	CTCGACAATC	GGACCATTTA	CACGCGCTGC	GTATCCGCCA	AAGGCAAATA
	501	CTACGCCACA	CCGCAAAAAC	TGAACGCCGA	TTACCATTTT	GAGCAAAGTG
50	551	TGCCTGCCGA	TATTTATTAC	ACGTTACTG	AAGAACATAC	CGACAAATCC
	601	AAGCTGTTG	CAATATCTT	ATATACGCC	CCCTTTTGA	TACTGGATGC
	651	GGCGGGCGCG	GTACTGGCCT	TGCCTGCGGC	GGCTCTGGGT	GCGGTCGTGG
	701	ATGCCGCCCG	CAATGA			

This corresponds to the amino acid sequence <SEQ ID 158; ORF28-1>:

55	1	MLFRKTTAAV	LAATLMLNGC	TLMLWGMNPN	VSETITRKHV	DKDQIRAFGV
	51	VAEDNAQLEK	GSLVMMGGKY	WFFVNPEDSA	KLTGILKAGL	DKPFQIVEDT
	101	PSYARHQALP	VKLESPGSQN	FSTEGLCRLY	DTDKPADIAK	LKQLGFCAVK
	151	LDNRTIYTRC	VSAKGKYYAT	PQKLNADYHF	EQSVPADIIY	TVTEHTDKS

201 KLFANILYTP PFLILDAAGA VLALPAAALG AVVDAARK\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF28 shows 79.2% identity over a 120aa overlap with an ORF (ORF28a) from strain A of *N.*

5 *meningitidis*:

```

10      orf28.pep      10      20      30      40      50      60
      MLFRKTTAAVLAHTLMLNGCTLMLWGMNPNVSETITRKHVXKDKQIRXFGVVAEDNAQLEK
      orf28a          10      20      30      40      50      60
      MLFRKTTAAVLAATLMLNGCTVMMWGMNSPFSETTARKHVDKDKQIRAFGVVAEDNAQLEK

15      orf28.pep      70      80      90      100     110     120
      GSLVMMGGKYWFVVPEDSAXXTGILKAGLDKPFQIVXDTPSYXCHQALPVKLGSGXSQN
      orf28a          70      80      90      100     110
      GSLVMMGGKYWFVVPEDSAKLTGILKAGLDKQFQMVPEPNRFA-YQALPVKLESPASQN

      orf28a          120     130     140     150     160     170
      FSTEGLCRLRYDTRDPADIAKLKQLEFEAVELDNRTIYTRCVSAKGKYYATPQKLNADYHF

```

20 The complete length ORF28a nucleotide sequence <SEQ ID 159> is:

```

25      1  ATGTTGTTCC  GTAAACGAC  CGCCGCCGTT  TTGGCGGCAA  CCTGATGTT
      51  GAACGGCTGT  ACGGTAATGA  TGTGGGGTAT  GAACAGCCCG  TTCAGCGAAA
      101  CGACCGCCCG  CAAACACGTT  GACAAGGACC  AAATCCGCGC  CTTCGGTGTG
      151  GTTGCCGAAG  ACAATGCCCA  ATTGGAAGAG  GGCAGCCTGG  TGATGATGGG
      201  CGGGAAATAC  TGGTTCGTCG  TCAATCTGA  AGATTCGGCG  AAGCTGACGG
      251  GCATTTTGAA  GGCCGGGTTG  GACAAGCAGT  TTCAAATGGT  TGAGCCCAAC
      301  CCGCGCTTTG  CCTACCAAGC  CCTGCCGGTC  AAACCTGAAT  CGCCCGCCAG
      351  CCAGAATTTC  AGTACCGAAG  GCCTTTGCCT  GCGCTACGAT  ACCGACAGAC
      401  CTGCCGACAT  CGCCAAGCTG  AAACAGCTTG  AGTTGAAGC  GGTGGAATC
      451  GACAATCGGA  CCATTTACAC  GCGCTGCGTC  TCCGCCAAAG  GCAAATACTA
      501  CGCCACACCG  CAAAACTGA  ACGCCGATTA  TCATTTTGAG  CAAAGTGTGC
      551  CTGCCGATAT  TTATTACACG  GTTACGAAAA  AACATACCGA  CAAATCCAAG
      601  TTGTTTGAAA  ATATTGCATA  TACGCCCACC  ACGTTGATAC  TGGATGCGGT
      651  GGGCGCGGTG  CTGGCCTTGC  CTGTCGCGGC  GTTGATTGCA  GCCACGAATT
      701  CCTCAGACAA  ATGA

```

This encodes a protein having amino acid sequence <SEQ ID 160>:

```

40      1  MLFRKTTAAV  LAATLMLNGC  TVMMWGMNSP  FSETTARKHV  DKDKQIRAFGV
      51  VAEDNAQLEK  GSLVMMGGKY  WVVVPEDSA  KLTGILKAGL  DKQFQMVPEPN
      101  PRFAYQALPV  KLESPASQNF  STEGLCLRYD  TDRPADIAKL  KQLEFEAVEL
      151  DNRTIYTRCV  SAKGKYYATP  QKLNADYHFE  QSVPADIIYT  VTKKHTDKSK
      201  LFENIAYTPT  TLILDAVGAV  LALPVAALIA  ATNSSDK*

```

ORF28a and ORF28-1 show 86.1% identity in 238 aa overlap:

```

45      orf28a.pep      10      20      30      40      50      60
      MLFRKTTAAVLAATLMLNGCTVMMWGMNSPFSETTARKHVDKDKQIRAFGVVAEDNAQLEK
      orf28-1          10      20      30      40      50      60
      MLFRKTTAAVLAATLMLNGCTLMLWGMNPNVSETITRKHVVDKDKQIRAFGVVAEDNAQLEK

50      orf28a.pep      70      80      90      100     110     119
      GSLVMMGGKYWFVVPEDSAKLTGILKAGLDKQFQMVPEPNRFA-YQALPVKLESPASQN
      orf28-1          70      80      90      100     110     120
      GSLVMMGGKYWFVVPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPASQN

55      orf28a.pep      120     130     140     150     160     170     179
      FSTEGLCRLRYDTRDPADIAKLKQLEFEAVELDNRTIYTRCVSAKGKYYATPQKLNADYHF
      orf28-1          120     130     140     150     160     170     180
      FSTEGLCRLRYDTRDPADIAKLKQLEFEAVELDNRTIYTRCVSAKGKYYATPQKLNADYHF

```

-142-

180 190 200 210 220 230  
 orf28a.pep EQSVPAADIYYTVTKKHTDKSKLFENIAYTPTTLILDVAGAVLALPVAALIAATNSSDKX  
 |||||:|||||:| |||||:|||||:| |||||:|  
 5 orf28-1 EQSVPAADIYYTVTEEHTDKSKLFANILYTPFLILDAAGAVLALPAAALGAVVDAARKX  
 190 200 210 220 230

Homology with a predicted ORF from *N.gonorrhoeae*ORF28 shows 84.2% identity over a 120aa overlap with a predicted ORF (ORF28.ng) from *N.*10 *gonorrhoeae*:

orf28.pep MLFRKTTAAVLAHTLMLNGCTLMLWGMNPNVSETITRKHVXKDQIRXFGVVAEDNAQLEK 60  
 |||||:|||||:| |||||:|||||:| |||||:|  
 orf28ng MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSETITRKHVDKDQIRAFGVVAEDNAQLEK 60  
 15 orf28.pep GSLVMMGGKYWFVNPEDSAXXTGILXAGLDKPFQIVXDTPSYXCHQALPVKLGSGXGSON 120  
 |||||:|||||:| |||||:|||||:| |||||:|  
 orf28ng GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSON 120

The complete length ORF28ng nucleotide sequence &lt;SEQ ID 161&gt; is

20 1 ATGTTGTTCC GTAAAACGAC CGCCGCCGTT TTGGCGGCAA CCTTGATACT  
 51 GAACGGCTGT ACGATGATGT TGCGGGGGAT GAACAACCCG GTCAGCCAAA  
 101 CAATCACCCG CAAACACGTT GACAAAGACC AAATCCGCGC CTTCCGTGTG  
 151 GTTGCCGAAG ACAATGCCCA ATTGGAAAAG GGCAGCCTGG TGATGATGGG  
 201 CGGGAAATAC TGGTTCGCCG TCAATCCCGA AGATTCGGCG AAGCTGACGG  
 251 GCCTTTTGAA GGCCGGGTTG GACAAGCCCT TCCAAATAGT TGAGGATACC  
 25 301 CCGAGCTATG CCCGCCACCA AGCCCTGCCG GTCAAATTCG AAGCGCCCGG  
 351 CAGCCAGAAT TTCAGTACCG GAGGTCTTTG CCTGCGCTAT GATACCGGCA  
 401 GACCTGACGA CATCGCCAAG CTGAAACAGC TTGAGTTTAA AGCGGTCAAA  
 451 CTCGACAATC GGACCATTTA CACGCGCTGC GTATCCGCCA AAGGCAAATA  
 501 CTACGCCACG CCGCAAAAAC TGAACGCCGA TTATCATTTT GAGCAAAGTG  
 30 551 TGCCCGCCGA TATTTATTAT ACGGTTACTG AAAAAACATC CGACAAATCC  
 601 AAGCTGTTTG GAAATATCTT ATATACGCCC CCCTTGTGTA TATTGGATGC  
 651 GCGCGCCGCG GTGCTGGTCT TGCCTATGGC TCTGATTGCA GCCGCGAATT  
 701 CCTCAGACAA ATGA

This encodes a protein having amino acid sequence &lt;SEQ ID 162&gt;:

35 1 MLFRKTTAAV LAATLILNGC TMMLRGMNPN VSQTITRKHV DKDQIRAFGV  
 51 VAEDNAQLEK GSLVMMGGKY WFAVNPEDSA KLTGLLKAGL DKPFQIVEDT  
 101 PSYARHQALP VKFEAPGSON FSTGGLCLRY DTGRPDIAK LKQLEFKAVK  
 151 LDNRTIYTRC VSAKGYYAT PQKLNADYHF EQSVPAADIY TVTEKHTDKS  
 201 KLFGNILYTP PLLILDAAAA VLVLPMALIA AANSSDK\*

40 ORF28ng and ORF28-1 share 90.0% identity in 231 aa overlap:

10 20 30 40 50 60  
 orf28-1.pep MLFRKTTAAVLAATLMLNGCTLMLWGMNPNVSETITRKHVDKDQIRAFGVVAEDNAQLEK  
 |||||:|||||:| |||||:|||||:| |||||:|  
 45 orf28ng MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSETITRKHVDKDQIRAFGVVAEDNAQLEK  
 10 20 30 40 50 60  
 70 80 90 100 110 120  
 orf28-1.pep GSLVMMGGKYWFVNPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPGSON  
 |||||:|||||:| |||||:|||||:| |||||:|  
 50 orf28ng GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSON  
 70 80 90 100 110 120  
 130 140 150 160 170 180  
 55 orf28-1.pep FSTGGLCLRYDTDKPADIAKLQLGFEAVKLDNRTIYTRCVSAKGYYATPQKLNADYHF  
 |||||:|||||:| |||||:|||||:| |||||:|  
 orf28ng FSTGGLCLRYDTGRPDIAKLQLEFAVKLDNRTIYTRCVSAKGYYATPQKLNADYHF  
 130 140 150 160 170 180  
 190 200 210 220 230 239  
 60 orf28-1.pep EQSVPAADIYYTVTEEHTDKSKLFANILYTPFLILDAAGAVLALPAAALGAVVDAARKX  
 |||||:|||||:| |||||:|||||:| |||||:|  
 orf28ng EQSVPAADIYYTVTEKHTDKSKLFGNILYTPPLLILDAAAALVLVLPMALIAAANSSDKX

190                      200                      210                      220                      230

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF28-1 (24kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 6A shows the results of affinity purification of the GST-fusion protein, and Figure 6B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result. These experiments confirm that ORF28-1 is a surface-exposed protein, and that it may be a useful immunogen.

### Example 20

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 163>:

```

15      1  ..GTCAGTCCTG TACTGCCTAT TACACACGAA CGGACAGGGT TTGAAGGTGT
      51  TATCGGTTAT GAAACCCATT TTTCAGGGCA CGGACATGAA GTACACAGTC
      101 CGTTCGATCA TCATGATTCA AAAAGCACTT CTGATTTCAG CGGCGGTGTA
      151 GACGGCGGTT TTA CTGTTTGA CCAACTTCAT CGAACATGGT CGGAAATCCA
      201 TCCGGAGGAT GAATATGACG GGCCGCAAGC AGCG.ATTAT CCGCCCCCGG
      251 GAGGAGCAAG GGATATATAC AGCTATTATG TCAAAGGAAC TTCAACAAAA
20      301 ACAAAGACTA GTATTGTCCC TCAAGCCCCA TTTTCAGACC GTTGGCTAGA
      351 AGAAAATGCC GGTGCCGCCT CTGGT..

```

This corresponds to the amino acid sequence <SEQ ID 164; ORF29>:

```

25      1  ..VSPVLPITHE RTGFEGVIGY ETHFSGHGHE VHSFPDHHDS KSTSDFSGGV
      51  DGGFTVYQLH RTWSEIHPED EYDGPQAAXY PPPGGARDIY SYVVKGTSTK
      101 TKT SIVPQAP FSDRWLEENA GAASG..

```

Further work revealed the complete nucleotide sequence <SEQ ID 165>:

```

30      1  ATGAATTTGC CTATTCAAAA ATTCATGATG CTGTTTGCAG CAGCAATATC
      51  GTTGCTGCAA ATCCCCATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC
      101 GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGTAA ATACCATCTG
      151 TTTGGTAATG CTCGCGGCAG TGTTAAAAAG CGGGTTTACG CCGTCCAGAC
      201 ATTTGATGCA ACTGCGGTCA GTCCTGTACT GCCTATTACA CACGAACGGA
      251 CAGGGTTTGA AGGTGTTATC GGTATGAAA CCCATTTTTC AGGGCACGGA
      301 CATGAAGTAC ACAGTCCGTT CGATCATCAT GATTCAAAAA GCACTTCTGA
      351 TTTTCAGCGGC GGTGTAGACG GCGGTTTTAC TGTTTACCAA CTTTCATCGAA
      401 CAGGGTCGGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC
      451 GATTATCCGC CCCCGGAGG AGCAAGGGAT ATATACAGCT ATTATGTCAA
      501 AGGAACCTTCA ACAAAAACAA AGACTAATAT TGTCCCTCAA GCCCATTTT
      551 CAGACCGTTG GCTAAAAGAA AATGCCGGTG CCGCCTCTGG TTTTTCAGC
      601 CGTGCGGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA
40      651 TTGGTGGGCT AACCGTATGG ATGATGTTTCG CCGCATCGTC CAAGGTGCGG
      701 TTAATCCTTT TTAATGGGT TTCAAGGAG TAGGGATTGG GGCAATTACA
      751 GACAGTGCAG TAAGCCCGGT CACAGATACA GCCGCGCAGC AGACTCTACA
      801 AGGTATTAAT GATTAGGAA AATTAAGTCC GGAAGCACAA CTTGCTGCCG
      851 CGAGCCTATT ACAGGACAGT GCTTTTGCAG TAAAAGACGG TATCAACTCT
45      901 GCCAAACAAT GGGCTGATGC CCATCCAAT ATAACAGCTA CTGCCCAAAC
      951 TGCCCTTTCC GCAGCAGAGG CCGCAGGTAC GGTTTGGAGA GGTAATAAAG
      1001 TAGAACTTAA CCCGACTAAA TGGGATTGGG TTAAAAATAC CGGTTATAAA
      1051 AAACCTGCTG CCCGCCATAT GCAGACTTTA GATGGGGAGA TGGCAGGTGG
      1101 GAATAAACCT ATTAAATCTT TACCAAACAG TGCCGCTGAA AAAAGAAAC
50      1151 AAAATTTTGA GAAGTTAAT AGTAACTGGA GTTCAGCAAG TTTTGATTCA

```

-144-

1201 GTGCACAAAA CACTAACTCC CAATGCACCT GGTATTTTAA GTCCTGATAA  
 1251 AGTTAAAACT CGATACACTA GTTTAGATGG AAAAATTACA ATTATAAAAG  
 1301 ATAACGAAAA CAACTATTTT AGAATCCATG ATAATTCACG AAAACAGTAT  
 1351 CTTGATTCAA ATGGTAATGC TGTGAAAACC GGTAAATTTAC AAGGTAAGCA  
 5 1401 AGCAAAAGAT TATTTACAAC AACAACTCA TATCAGGAAC TTAGACAAAT  
 1451 GA

This corresponds to the amino acid sequence <SEQ ID 166; ORF29-1>:

1 MNLPIQKFM LFAAAISLLQ IPISHANGLD ARLRDDMQAK HYEPPGKYHL  
 51 FGNARGSVKK RVYAVQTFDA TAVSPVLPIT HERTGFEGVI GYETHFSGHG  
 10 101 HEVHSPFDHH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS  
 151 DYPPPGGARD IYSYVKGTS TKTKTNIVPQ APFSDRWLKE NAGAASGFFS  
 201 RADEAGKLIW ESDPNKNWWA NRMDDVRGIV QGAVNPFLMG FQGVGIGAIT  
 251 DSAVSPVTD TAAQTLQGIN DLGKLSPEAQ LAAASLLQDS AFAVKDGINS  
 301 AKQWADAHNP ITATAQTALS AAEAAGTVWR GKKVELNPTK WDWVKNVTGYK  
 15 351 KPAARHMQTL DGEMAGGNKP IKSLPNSAAE KRKQNFEEK SNWSSASFDS  
 401 VHKTLPNAP GILSPDKVKT RYTSLDGKIT IKDNENNYF RIHDNSRKQY  
 451 LDSNGNAVKT GNLQKGQAKD YLQQQTHIRN LDK\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF29 shows 88.0% identity over a 125aa overlap with an ORF (ORF29a) from strain A of *N. meningitidis*:

	orf29.pep				10	20	30	
					VSPVLPIT	HERTGFEGVI	GYETHFSGHGHE	
					: :	: :	: :	
25	orf29a	EPGGKYHLFGNARGSVK	NRVYAVQTFDATA	VGPILPIT	HERTGFEGVI	GYETHFSGHGHE		
		50	60	70	80	90	100	
	orf29.pep		40	50	60	70	80	90
			VHSPFDHHD	SKSTSDFS	GVDGGFTVYQL	HRTWSEIHP	EDGYDGPQA	XYPPPGGARDIY
30			: :	: :	: :	: :	: :	: :
	orf29a		VHSPFDNHD	SKSTSDFS	GVDGGFTVYQL	HRTGSEIHP	EDGYDGPQGS	DYPPPGGARDIY
			110	120	130	140	150	160
35	orf29.pep		100	110	120			
			SYVVGKTST	TKTSIVPQ	APFSDRWLE	NAGAASG		
			: :	: :	: :	: :	: :	: :
	orf29a		XXYVVGKT	STKTSNIV	PRAPFSDRW	LKENAGAAS	GFFSRADEAG	KLIWESDPN
			170	180	190	200	210	220
40	orf29a		MDDIRGIVQ	GAVNPFLMG	FQGVGIGAIT	DSAVSPVTD	TAAQTLQGX	NHLGXLSPEA
			230	240	250	260	270	280

The complete length ORF29a nucleotide sequence <SEQ ID 167> is:

1 ATGAATTNGC CTATTCAAAA ATTCATGATG CTGTTTGCGAG CAGCAATATC  
 45 51 GTNGCTGCAA ATCCCNATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC  
 101 GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGTAA ATACCATCTG  
 151 TTTGGTAATG CTCGCGGCAG TGTAAAAAT CGGGTTTACG CCGTCCAAAC  
 201 ATTTGATGCA ACTGCGGTCTG GCCCATACT GCCTATTACA CACGAACGGA  
 251 CAGGATTTGA AGGCATTATC GGTATGAAA CCCATTTTTC AGGACATGGA  
 301 CATGAAGTAC ACAGTCCGTT CGATAATCAT GATTCAAAAA GCACTTCTGA  
 50 351 TTTAGCGCGC GCGTAGACG GTGGTTTTAC CGTTTACCAA CTTTCATCGA  
 401 CAGGGTCGGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC  
 451 GATTATCCGC CCCCGGAGG AGCAAGGGAT ATATACANNT ANTATGTCAA  
 501 AGGAACCTCA ACAAAAACAA AGAGTAATAT TGTTCCTCCG GCGCATTTT  
 55 551 CAGACCGCTG GCTAAAAGAA AATGCCGGTG CCGCCTCTGG TTTTTCAGC  
 601 CGTGCTGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA  
 651 TTTGGTGGCT AACCGTATGG ATGATATTCG CGGCATCGTC CAAGGTGCGG  
 701 TTAATCCTTT TTTAATGGGT TTTCAAGGAG TAGGGATTGG GGCAATTACA  
 751 GACAGTGCAG TAAGCCCGGT CACAGATACA GCCGCGCAGC AGACTCTACA  
 801 AGGTATNAAT CATTTAGGAA ANTTAAGTCC CGAAGCACAA CTTGCGGCTG  
 60 851 CAACCGCATT ACAGACAGT GCTTTTGCAG TAAAAGACGG TATCAATTCC  
 901 GCCAGACAAT GGGCTGATGC CCATCCGAAT ATAAGTCAA CAGCCCAAC

-145-

951 TGCCCTTGCC GTAGCAGANG CCGCAACTAC GGTTTGGGGC GGTAAAAAAG  
 1001 TAGAACTTAA CCCGACCAAA TGGGATTGGG TAAAAAATAC NGGCTATAAN  
 1051 ACACCTGCTG TTCGCACCAT GCATACTTTG GATGGGGAAA TGGCCCGTGG  
 1101 GAATAGACCG CCTAAATCTA TAACGTCCAA CAGCAAAGCA GATGCTTCCA  
 1151 CACAACCGTC TTTACAAGCG CAACTAATTG GAGAACAAAT TANNNNNGGG  
 1201 CATGCTTATA ACAAGCATGT CATAAGACAA CAAGAATTTA CGGATTAAAA  
 1251 TATCAATTCA CCAGCAGATT TTGCTCGGCA TATTGAAAAT ATTGTTAGCC  
 1301 ATCCANCAAA TATGAAAGAG TTACCTCGCG GTAGAACTGC GTATTGGGAT  
 1351 NATAAAACAG GGACNATAGT TATCCGAGAT AAAAATTCTG ACGATGGAGG  
 1401 TACAGCATTT AGACCAACAT CAGGTAAAAA ATATTATGAT GATTATATAG

This encodes a protein having amino acid sequence <SEQ ID 168>:

1 MNXPIQKFM LFAAAISXLQ IPISHANGLD ARLRDDMQAK HYEPPGKYHL  
 51 FGNARGSVKN RVYAVQTFDA TAVGPILPIT HERTGFEGII GYETHFSGHG  
 101 HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS  
 151 DYPPPGGARD IYXXYVKGTS TKTKSNIVPR APFSDRWLKE NAGAASGFFS  
 201 RADEAGKLIW ESDPNKNWWA NRMDDIRGIV QGAVNPFLMG FQGVGIGAIT  
 251 DSAVSPVTD TAAQOTLQGXN HLGXLSPEAQ LAAATALQDS AFAVKDGIN  
 301 ARQWADAHNP ITATAQTALA VAXAATTVWG GKKVELNPTK WDWKNTGYX  
 351 TPAVRTMHTL DGEMAGGNRP PKSITSNSKA DASTQPSLQA QLIGEIXXG  
 401 HAYNKHVIRQ QEFTDLNINS PADFARHIEN IVSHPXNMKE LPRGRYAYWD  
 451 XKTGTIVIRD KNSDDGGTAF RPTSGKKYYD DL\*

ORF29a and ORF29-1 show 90.1% identity in 385 aa overlap:

		10	20	30	40	50	60
25	orf29a.pep	MNXPIQKFM LFAAAISXLQIPISHANGLDARLRDDMQAKHYEPPGKYHLFGNARGSVKN					
	orf29-1	MNLPIQKFM LFAAAISLLQIPISHANGLDARLRDDMQAKHYEPPGKYHLFGNARGSVKK					
		10	20	30	40	50	60
30	orf29a.pep	RVYAVQTFD TAVGPILPITHERTGFEGII GYETHFSGHGHEVHSPFDNH DSKSTSDFSG					
	orf29-1	RVYAVQTFD TAVSPVLPITHERTGFEGVIGYETHFSGHGHEVHSPFDH DSKSTSDFSG					
		70	80	90	100	110	120
35	orf29a.pep	GVDGGFTVYQLHRTGSEIHPEDGYDGPQGS DYPPPGGARDIYXXYVKGTSTKTKSNIVPR					
	orf29-1	GVDGGFTVYQLHRTGSEIHPEDGYDGPQGS DYPPPGGARDIYSYVKGSTKTKTNIVPQ					
		130	140	150	160	170	180
40	orf29a.pep	APFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANRMDDIRGIVQGAVNPFLMG					
	orf29-1	APFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANRMDDVRGIVQGAVNPFLMG					
		190	200	210	220	230	240
45	orf29a.pep	FQGVGIGAITDSAVSPVTD TAAQOTLQGXNHLGXLSPEAQ LAAATALQDS AFAVKDGIN					
	orf29-1	FQGVGIGAITDSAVSPVTD TAAQOTLQGINDLGKLSPEAQ LAAASLLQDS AFAVKDGIN					
		250	260	270	280	290	300
50	orf29a.pep	ARQWADAHNPITATAQTALAVAXAATTVWG GKKVELNPTKWDWKNTGYXTPAVRTMHTL					
	orf29-1	AKQWADAHNPITATAQTALSAEAAAGTVWRGKKVELNPTKWDWKNTGYKKPAARHMOTL					
		310	320	330	340	350	360
55	orf29a.pep	DGEMAGGNRPKPKSITSNSKADASTQPSLQAQLIGEIXXGHAYNKHVIRQQEFTDLNINS					
	orf29-1	DGEMAGGNKPIKSLP-NSAAEKRKQNF EKFSNSWSSASFDSVHKTLTPNAPGILSPDKVK					
		370	380	390	400	410	
60	orf29a.pep						
	orf29-1						

Homology with a predicted ORF from *N.gonorrhoeae*

ORF29 shows 88.8% identity over a 125aa overlap with a predicted ORF (ORF29.ng) from *N.*

*gonorrhoeae*:

5	orf29.pep	VSPVLPITHERTGFEGVIGYETHFSGHGHE	30
		:	
	orf29ng	EPGGKYHLFGNARGSVKNRVCVQTFDATAVGPILPITHERTGFEGVIGYETHFSGHGHE	102
	orf29.pep	VHSPFDHSDSKSTSDFSGGVDGGFTVYQLHRTWSEIHPEDGYDGPQAAXYPPPGGARDIY	90
10		:	
	orf29ng	VHSPFDNHSDSKSTSDFSGGVDGGFTVYQLHRTGSEIHPEDGYDGPQGGYPPPGGARDIY	162
	orf29.pep	SYVVKGTSTKTTSIVPQAPFSDRWLEENAGAASG	125
		:      :	
	orf29ng	SYHIKGTSTKTINTVPQAPFSDRWLKENAGAASGFLSRADEAGKLIWENDPKNWRANR	222

15 The complete length ORF29ng nucleotide sequence <SEQ ID 169> is predicted to encode a protein having amino acid sequence <SEQ ID 170>:

	1	MNLPIQKFMM	LFAAAISLLQ	IPISHANGLD	ARLRDDMQAK	HYEPGGKYHL
	51	FGNARGSVKN	RVCVQTFDA	TAVGPILPIT	HERTGFEGVI	GYETHFSGHG
20	101	HEVHSPFDNH	DSKSTSDFSG	GVDGGFTVYQ	LHRTGSEIHP	EDGYDGPQGG
	151	GYPPPGGARD	IYSYHIKGT	TKTKINTVPQ	APFSDRWLKE	NAGAASGFLS
	201	RADEAGKLIW	ENDDPKNWRA	NRMDDIRGIV	QGA VNPFLT	FGGLGVGAIT
	251	DSAVSPVTYA	AARKTLQGIH	NLGNLSPEAQ	LAAATALQDS	AFAVKDSINS
	301	ARQWADAHPN	ITATAQTALA	VTEAATTWVG	GKKVELNPAK	WDVWKNTGYK
	351	KPAARHMOTV	DGEMAGGNKP	LESKNTVTTN	NFFENTGYTE	KVLRQASNGD
25	401	YHGFPQSVDA	FSENGTVIQI	YHGDNI VRHK	LYIPGSYK GK	DGNFEYIREA
	451	DGKINHRLFV	PNQQLPEK*			

In a second experiment, the following DNA sequence <SEQ ID 171> was identified:

	1	atgAATTTGC	CTATTCAAAA	ATTCATGATG	ctgttggcAg	cggcaatatc
	51	gatgctGCat	ATCCCCATTA	GTCATGCGAA	CGGTTTGGAT	GCCCCGTTTGC
30	101	GCGATGATAT	GCAGGCAAAA	CACTACGAAC	CGGGTGGCAA	ATACCATCTG
	151	TTTGTTAATG	CTCGCGGCAG	TGTTAAAAAT	CGGGTTTGCG	CCGTCCAAAC
	201	ATTTGATGCA	ACTGCGGTCG	GCCCCATACT	GCCTATTACA	CACGAACGGA
	251	CAGGATTTGA	AGGTGTTATC	GGCTATGAAA	CCCATTTTTC	AGGACACGGA
	301	CACGAAGTAC	ACAGTCCGTT	CGATAATCAT	GATTCAAAAA	GCACTTCTGA
35	351	TTTCAGCGGC	GGCGTAGACG	GCGGTTTTAC	CGTTTACCAA	CTTCATCGGA
	401	CAGGGTCGGA	AATACATCCC	GCAGACGGAT	ATGACGGGCC	TCAAGGCGGC
	451	GGTTATCCGG	AACCACAAGG	GGCAAGGGAT	ATATACAGCT	ACCATATCAA
	501	AGGAACTTCA	ACCAAAACAA	AGATAAACAC	TGTTCCGCAA	GCCCCTTTTT
	551	CAGACCGCTG	GCTAAAAGAA	AATGCCGGTG	CCGCTTCCGG	TTTTCTCAGC
40	601	CGTGCGGATG	AAGCAGGAAA	ACTGATATGG	GAAAACGACC	CCGATAAAAA
	651	TTGGCGGGCT	AACCGTATGG	ATGATATTCG	CGGCATCGTC	CAAGGTGCGG
	701	TTAATCCTTT	TTTAACGGGT	TTTCAAGGGG	TAGGGATTGG	GGCAATTACA
	751	GACAGTGCGG	TAAGCCCGGT	CACAGATACA	GCCGCTCAGC	AGACTCTACA
45	801	AGGTATTAAT	GATTTAGGAA	ATTTAAGTCC	GGAAGCACAA	CTTGCCGCCG
	851	CGAGCCTATT	ACAGGACAGT	GCCTTTGCGG	TAAAAGACGG	CATCAATTCC
	901	GCCAGACAAT	GGGCTGATGC	CCATCCGAAT	ATAACAGCAA	CAGCCCAAAC
	951	TGCCCTTGCC	GTAGCAGAGG	CCGCAGGTAC	GGTTTGGCGC	GGTAAAAAAG
	1001	TAGAACTTAA	CCCCACCAAA	TGGGATTGGG	TAAAAAATAC	CGGCTATAAA
	1051	AAACCTGCTG	CCCGCCATAT	GCAGACTGTA	GATGGGGAGA	TGGCAGGGGG
50	1101	GAATAGACCG	CCTAAATCTA	TAACGTCGGA	AGGAAAAGCT	AATGCTGCAA
	1151	CCTATCCTAA	GTTGGTTAAT	CAGCTAAATG	AGCAAACTT	AAATAACATT
	1201	GCGGCTCAAG	ATCCAAGATT	GAGTCTAGCT	ATTCATGAGG	GTAAAAAATA
	1251	TTTTCCTAAT	GGAAGTCAA	CTTATGAAGA	GGCAGATAGA	CTAGGTAAAA
	1301	TTTGGGTTGG	TGAGGTGCA	AGACAACTA	GTGGAGGCGG	ATGGTTAAGT
55	1351	AGAGATGGCA	CTCGACAATA	TCGGCCACCA	ACAGAAAAAA	AATCACAATT
	1401	TGCAACTACA	GGTATTCAAG	CAAATTTTGA	AACTTATACT	ATTGATTCAA
	1451	ATGAAAAAAG	AAATAAAATT	AAAAATGGAC	ATTTAAATAT	TAGGTAA

This encodes a protein having amino acid sequence <SEQ ID 172; ORF29ng-1>:

60	1	MNLPIQKFMM	LLAAAISMLH	IPISHANGLD	ARLRDDMQAK	HYEPGGKYHL
	51	FGNARGSVKN	RVCVQTFDA	TAVGPILPIT	HERTGFEGVI	GYETHFSGHG



ORF29ng-1 and ORF29-1 show 86.0% identity in 401 aa overlap:

60

### Example 21

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 173>:

-148-

```

1  ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
51  CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAATGTTCC
101 ACACGCGGGC AGATGCACCG ATGCAG...

```

This corresponds to the amino acid sequence <SEQ ID 174; ORF30>:

5 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QMFHTRADAP MQ..

Further work revealed the complete nucleotide sequence <SEQ ID 175>:

```

1  ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
51  CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTC
101 ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG
10  151 ATGAAGGAGA CAGAGGGGGC GTTCTTCCA TTGGCTATCT TGGGTGGTGC
201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
251 GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
301 CCTGGTGGTG TAGGCGCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG
351 ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCTT TCGGTAATA
15  401 GAACAGGTCA TCCTATTGGA AAATTTCCCC ATTATCATCG TCGAGTTACG
451 GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
501 TTGGGAATCA AAATCTACGG ACAGATCATG GAAAACCGC TTCTAA

```

This corresponds to the amino acid sequence <SEQ ID 176; ORF30-1>:

```

20 1  MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
51  MKETEGAFLP LAILGGAAIG MWTQHGFSA TTGRPASVRD VAIAGGLGAI
101 PGGVGAAGKV VSAKYGREI KIGNNMRIAP FGNRTGHPIG KFPHYHRRVT
151 DNTGKTLPGQ GIGRHRPWES KSTDRSWKNR F*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF30 shows 97.6% identity over a 42aa overlap with an ORF (ORF30a) from strain A of *N. meningitidis*:

```

30 orf30.pep      10      20      30      40
      MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ
      |||||:|||||
orf30a      MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP
      10      20      30      40      50      60
orf30a      LXILGGAAIGMWTQHGFSAATTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI
      70      80      90      100     110     120

```

35 The complete length ORF30a nucleotide sequence <SEQ ID 177> is:

```

1  ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
51  CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTC
101 ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG
40  151 ATGAAGGANA CAGNGGGGGC GTTCTTCCA TTGNTATCT TGGGTGGTGC
201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
251 GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
301 CCTGGTGN TG TAGGCGCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG
351 ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCTT TCGGTAATA
40  401 GAACAGGTCA TCCTATTGGA AAATTTCCCC ATTATCATCG TCGAGTTACG
45  451 GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
501 TTGGGAATCA AAATCTACGG ACAGATCATG GAAAACCGC TTCTAA

```

This encodes a protein having amino acid sequence <SEQ ID 178>:

```

50 1  MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
51  MKXTXGAFLP LXILGGAAIG MWTQHGFSA TTGRPASVRD VAIAGGLGAI
101 PGXVGAAGKV VSAKYGREI KIGNNMRIAP FGNRTGHPIG KFPHYHRRVT
151 DNTGKTLPGQ GIGRHRPWES KSTDRSWKNR F*

```

ORF30a and ORF30-1 show 97.8% identity in 181 aa overlap:

```

orf30a.pep      MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP 60

```

	orff30-1	 MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP	60
5	orff30a.pep	LXILGGAAIGMWTQHGFYSYATTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI	120
	orff30-1	 LAILGGAAIGMWTQHGFYSYATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI	120
	orff30a.pep	KIGNNMRIAPFGNRTGHPIGKFPYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR	180
10	orff30-1	 KIGNNMRIAPFGNRTGHPIGKFPYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR	180
	orff30a.pep	FX	
15	orff30-1	FX	

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF30 shows 97.6% identity over a 42aa overlap with a predicted ORF (ORF30.ng) from *N. gonorrhoeae*:

20	orff30.pep	MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ	42
	orff30ng	MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP	60

The complete length ORF30ng nucleotide sequence <SEQ ID 179> is

	1	ATGAAAAAAC	AAATCACCGC	AGCCGTAATG	ATGCTGTCTA	TGATCGCCCC
25	51	CGCAATGGCA	AACGGATTGG	ACAATCAGGC	ATTGAAGAC	CAAGTGTTC
	101	ACACGCGGGC	AGATGCGCCG	ATGCAGTTGG	CGGAGCTTTC	TCAGAAGGAG
	151	ATGAAGGAGA	CTGAAGGGGC	TTTTCTTCCA	TTGGCTATCT	TGGGTGGTGC
	201	TGCCATTGGT	ATGTGGACAC	AGCATGTTT	TAGTTATGCA	ACGACAGGCA
	251	GACCAGCTTC	TGTTAGAGAT	GTTGCTGGCG	GATTAGGCGC	AATTCCTGGT
30	301	GATGTAGGTG	CTGCAGGAAA	GGTTGTTTCC	TTTGCTAAAT	ATGACGTGA
	351	GATTAAAATC	GGCAATAATA	TGCGGATAGC	CCCTTTCGGT	AATAGAACAG
	401	GTCATCCTAT	TGGAAAATTT	CCCCATTATC	ATCGTCGAGT	TACGGATAAT
	451	ACGGGCAAGA	CTTTGCCTGG	ACAGGGAATT	GGTCGTCATC	GCCCTTGGGA
	501	ATCAAAATCT	ACGGACAGAT	CATGGAAAAA	CCGCTTCTAA	

This encodes a protein having amino acid sequence <SEQ ID 180>:

35	1	MKKQITAAVM	MLSMIAPAMA	NGLDNQAFED	QVFHTRADAP	MQLAELSQKE
	51	MKETEGAFLP	LAILGGAAIG	MWTQHGFSYA	TTGRPASVRD	VAGGLGAIPG
	101	DVGAAGKVVS	FAKYGREIKI	GNNMRIAPFG	NRTGHPIGKF	PHYHRRVTDN
	151	TGKTLPGQGI	GRHRPWESKS	TDRSWKNRF*		

ORF30ng and ORF30-1 show 98.3% identity in 181 aa overlap:

40		10	20	30	40	50	60
	orff30ng.pep	MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP					
	orff30-1						
45		10	20	30	40	50	60
	orff30ng.pep	LAILGGAAIGMWTQHGFYSYATTGRPASVRDVA--GGLGAIPGDVGAAGKVVSFAKYGREI					
	orff30-1						
50		70	80	90	100	110	120
	orff30ng.pep	LAILGGAAIGMWTQHGFYSYATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI					
	orff30-1						
55		120	130	140	150	160	170
	orff30ng.pep	KIGNNMRIAPFGNRTGHPIGKFPYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR					
	orff30-1						
		130	140	150	160	170	180
60	orff30ng.pep	FX					
	orff30-1	FX					

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 22

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 181>:

10

1	ATGAATAAAA	CTCTCTATCG	TGTAATTTTC	AACCGCAAAC	GTGGGGCTGT
51	GrTAGCCGTT	GCTGAAACTA	CCAAGCGCGA	AGGTAAAGC	TGTGCCGATA
101	GTGATTCAAG	CAGCGCTCAT	GTGAAATCTG	TTCTTTTGG	TACTACTCAT
151	GCACCTGTTT	GTg.CGTTaC	AAATATCTTT	TCTTTTCTT	TATTGGGCTT
201	TTCTTTTATGT	TTGGCTGTAG	GtacGGyCAA	TATTGCTTTT	GCTGATGGCA
251	TT...				

This corresponds to the amino acid sequence <SEQ ID 182; ORF31>:

1 MNKTLYRVIF NRKRGAVXAV AETTKREGKS CADSDSGSAH VKSVFPGTTH  
51 APVCXVTNIF SFSLLGFSLC LAVGTXNIAF ADGI..

15 Further work revealed a further partial nucleotide sequence <SEQ ID 183>:

20

1	ATGAATAAAA	CTCTCTATCG	TGTAATTTTC	AACCGCAAAC	GTGGGGCTGT
51	GGTAGCCGTT	GCTGAAACTA	CCAAGCGCGA	AGGTAAAAGC	TGTGCCGATA
101	GTGATTACAG	CAGCGCTCAT	GTGAAATCTG	TTCCTTTTGG	TACTACTCAT
151	GCACCTGTTT	GTCGTTCAA	TATCTTTTCT	TTTTCTTTAT	TGGGCTTTTC
201	TTTATGTTTG	GCTGTAGGTA	CGGCCAATAT	TGCTTTTGCT	GATGGCATT.

This corresponds to the amino acid sequence <SEQ ID 184; ORF31-1>:

1 MNKTLRVIF NRKRGAVVAV AETTKREGKS CADSDSGSAH VKSVFPGTTH  
51 APVCRSNIFF FSLLGFSLCL AVGTANIAFA DGI..

Computer analysis of this amino acid sequence gave the following results:

25 Homology with a predicted ORF from *N.gonorrhoeae*

ORF31 shows 76.2% identity over a 84aa overlap with a predicted ORF (ORF31.ng) from *N. gonorrhoeae*:

	orf31.pep	MNKTLYRVIFNRKRGA <del>V</del> XA <del>V</del> AE <del>T</del> TKREGKSCADSDSGSAHVKSVPFGTTHAPVCXVTNIF	60
30			
	orf31ng	MNKTLYRVIFNRKRGA <del>V</del> V <del>A</del> AE <del>T</del> TKREGKSCADSDSGSVYVKSVSFIPTH-----SKAF	54
	orf31.pep	SFSLGLGFSLCLAVGT <del>X</del> NIAFADGI	84
	orf31ng	CFSALGFSLCLALGTVNIAFADGIITDKAAPK <del>T</del> QOATILQ <del>T</del> NGNIPQVNIQTPTSAGVSV	114

35 The complete length ORF31ng nucleotide sequence <SEQ ID 185> is:

	1	ATGAACAAAA	CCCTCTATCG	TGTGATTTTC	AACCGCAAAC	GCGGTGCTGT
	51	GGTAGCTGTT	GCCGAAACCA	CCAAGCGCGA	AGGTAAAAGC	TGTGCCGATA
	101	GTGGTTCGGG	CAGCGTTTAT	GTGAAATCCG	TTTCTTTTCAT	TCCTACTCAT
40	151	CTCAAAGCCT	TTTGTTTTTC	TGCATTAGGC	TTTTCTTTAT	GTTTGGCTTT
	201	GGTACGGTC	AATATTGCTT	TTGCTGACGG	CATTATTACT	GATAAAGCTG
	251	CTCCTAAAAC	CCAACAAGCC	ACGATTCTGC	AAACAGGTaa	cGGCATAACG
	301	CAAGTCAATA	TTCAAACCCC	TACTTCGGCA	GGGGTTTTCTG	TTAATCAATA
	351	TGCCCAGTTT	GATGTGGGTA	ATCGCGGGGC	GATTTTAAAC	AACAGTCGCA
45	401	GCAACACCCA	AACACAGCTA	GCGGTTGGA	TTCAAGGCaa	TCCTTGGTTG
	451	ACAAGGGGCG	AAGCACGTGT	GGTGTGAAAC	CAAATCAACA	GCAGCCATCC
	501	TTCACAACTG	AATGGCTATA	TTGAAGTGGG	TGGACGACGT	GCAGAAAGTCG
	551	TTATTGCCAA	TCCGGCAGGG	ATTGCAGTCA	ATGGTGGTGG	TTTTTATCAAT
	601	GCTTCCCCTG	CCACTTTGAC	GACAGGCCAA	CCGCAATATC	AAGCAGGAGA
	651	CTTTAGCGGC	TTTAAGATAA	GGCAAGGCCAA	TGCTGTAATC	GCCGGACACG

-151-

701 GTTGGATGC CCGTGATACC GATTTCACAC GTATTCTTGT ATGCCAACAA  
 751 AATCACCTTG ATCAGTACGG CCGAACAAAGC AGGCATTCGT AA

This encodes a protein having amino acid sequence <SEQ ID 186>:

5 1 MNKTLRVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIPTH  
 51 SKAFCFSALG FSLCLALGTV NIAFADGIIT DKAAPKTQQA TILQTGNGIP  
 101 QVNIQTPTSA GVSVNQYAQF DVGNRGAILN NSRSNTQTQL GGWIQGNPWL  
 151 TRGEARVVVN QINSSHPSQL NGYIEVGRR AEVVIANPAG IAVNGGGFIN  
 201 ASRATLTGQ PQYQAGDFSG FKIRQNAVI AGHGLDARDT DFTRILVCQQ  
 251 NHDQYGRS RHS\*

10 This gonococcal protein shares 50% identity over a 149aa overlap with the pore-forming hemolysins-like HecA protein from *Erwinia chrysanthemi* (accession number L39897):

15 orf31ng 96 GNGIPQVNIQTPTSAGVSVNQAQFDVGNRGAILNNSRSN-TQTQLGGWIQGNPWLTRGE 154  
 GNG+P VNI TP ++G+S N+Y F+V NRG ILNN + T +QLGG IQ NP L  
 HecA 45 GNGVPPVNIATPDASGLSHNRYHDFVNDNRGLILNNGTARLTGPSQLGGLIQNNPNLNGRA 104  
 20 Orf31ng 155 ARVVVNQINSSHPSQLNGYIEVGRRRAEVVIANPAGIAVNGGGFINASRATLTGQPQYQ 214  
 A ++N++ S + S+L GY+EV G+ A VV+ANP GI +G GF+N R TLTG PQ+  
 HecA 105 AAAILNEVVSPNRSRLAGYLEVAGQAANVVVANPYGITCSGCGFLNTPRLTLTTGTGPQFD 164  
 20 Orf31ng 215 -AGDFSGFKIRQGNNAVIAGHGLDARDTDF 242  
 AG SG +R G+ +I G GLDA +D+  
 HecA 165 AAGGLSGLDVRGGDILIDGAGLDASRSY 193

Furthermore, ORF31ng and ORF31-1 show 79.5% identity in 83 aa overlap:

25 orf31-1.pep 10 20 30 40 50 60  
 MNKTLRVIFNRKRGAVVAVAEETTKREGKSCADSDSGSAHVKSVPFGTTHAPVCRSNIFS  
 orf31ng 10 20 30 40 50  
 MNKTLRVIFNRKRGAVVAVAEETTKREGKSCADSGSGSVYVKSVSFIPTH-----SKAFC  
 30 orf31-1.pep 70 80  
 FSLLGFSCLAVGTANIAFADGI  
 orf31ng 60 70 80 90 100 110  
 FSALGFSCLALGTVNIAFADGIITDKAAPKTQQAATILQTGNGIPQVNIQTPTSAGVSVN

35 On this basis, including the homology with hemolysins, and also with adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 23

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 187>:

40 1 ATGAATACTC CTCCTTTTGT CTGTTGGATT TTTTGCAAGG TCATCGACAA  
 51 TTTCGGCGAC ATCGGCGTTT CGTGGCGGCT CGCCCGTGT TTGCACCGCG  
 101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT  
 151 GCGCTTTGCC CTGATTGCC CGATGTTCCC TGCCTTCATC AGGATATTCA  
 201 TGTCGCACT TGCAATCCG ATGCGGCAGA TATTGATACC GCG..

45 This corresponds to the amino acid sequence <SEQ ID 188; ORF32>:

1 MNTPPFVCWI FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR  
 51 ALCPDLPDVP CVHQDIHVRT WHSDAADIDT A..

Further work revealed the complete nucleotide sequence <SEQ ID 189>:

50 1 ATGAATACTC CTCCTTTTGT CTGTTGGATT TTTTGCAAGG TCATCGACAA  
 51 TTTCGGCGAC ATCGGCGTTT CGTGGCGGCT CGCCCGTGT TTGCACCGCG  
 101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT

-152-

5  
10  
15  
20

```

151 GCGCTTTGCC CTGATTTGCC CGATGTTCCC TCGGTTTCATC AGGATATTCA
201 TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTC
251 CCGATGTCGT CATCGAAACT TTTGCCTGCG ACCTGCCCGA AAATGTGCTG
301 CACATTATCC GCCGACACAA GCCGCTTTGG CTGAATTGGG AATATTTGAG
351 CGCGGAGGAA AGCAATGAAA GGCTGCATCT GATGCCTTCG CCGCAGGAGG
401 GTGTTCAAAA ATATTTTGGG TTTATGGGTT TCAGCGAAAA AAGCGGCGGG
451 TTGATACGCG AACGTGATTA CTGCGAAGCC GTCCGTTTCG ATACTGAAGC
501 CCTGCGAGAG CGGCTGATGC TGCCCGAAAA AAACGCCTCC GAATGGCTGC
551 TTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
601 CAGGCAGGCA GCCCGATGAC ACTGTTGCTG GCGGGGACGC AAATCATCGA
651 CAGCCTCAAA CAAAGCGGCG TTATTCCGCA AGATGCCCTG CAAAACGACG
701 GCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG
751 CCGCAACAGG ACTTCGACCA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT
801 CCGCGGCGAA GACAGTTTCG TGCGCGCCCA GCTTGCGGGC AAACCTTCT
851 TTTGGCACAT CTACCCGCAA GACGAGAATG TCCATCTCGA CAAACTCCAC
901 GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGTGTCCGC
951 ACACCGCCGT CTTTCGGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
1001 CACAACGCCT CGAATGTTGG CAAACCTGCA AACAACATCA AAACGGCTGG
1051 CGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTCGGGC AGCCGTCAGC
1101 TCCTGAAAAA CTCGCTGCCT TTGTTTCAAA GCATCAAAA ATACGCTAG

```

This corresponds to the amino acid sequence <SEQ ID 190; ORF32-1>:

25  
30

```

1 MNTPPFVCWI FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
51 ALCPDLPDVP CVHQDIHVRT WHSDAADIDT APVPDVVIET FACDLPENVL
101 HIIRRHKPLW LNWEYLSAEE SNERLHLMPS PQEGVQKYFW FMGFSEKSGG
151 LIRERDYCEA VREFTEALRE RLMLPEKNAS EWLLFGYRSD VWAKWLEMR
201 QAGSPMTLLL AGTQIIDSLE QSGVIPQDAL QNDGDVFQTA SVRLVKIPFV
251 PQQDFDQLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPO DENVHLDKLH
301 AFWDKAHGFY TPETVSAHRR LSDDLNGGEA LSATORLECW QTLQHQHNGW
351 RQGAEDWSRY LFGQPSAPEK LAAFVSKHQK IR*w

```

30 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF32 shows 93.8% identity over a 81aa overlap with an ORF (ORF32a) from strain A of *N.*

*meningitidis*:

35  
40

```

          10      20      30      40      50      60
orf32.pep MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVP
          |||||  |||||||  |||||||  |||||||  |||||||  |||||||
orf32a    MNTPPFSAGXFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX
          10      20      30      40      50      60

          70      80
orf32.pep CVHQDIHVRTWHSDAADIDTA
          |||||  |||||  |||||  |||||  |||||  |||||
orf32a    CVHQDIHVRTWHSDAADIDTAPVXDVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX
          70      80      90      100     110     120

```

45 The complete length ORF32a nucleotide sequence <SEQ ID 191> is:

50  
55  
60

```

1 ATGAATACTC CTCCTTTTTC TGCTGGANTT TTTTGCAAGG TCATCGACAA
51 TTTTCGGCGAC ATCGGCGTTT CGTGGCGGCT TGCCCGTGT TTTGACCGCG
101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT
151 GCGCTTTGCC CTGATTTGCC CGATGTTTNC TCGGTTTCATC AGGATATTCA
201 TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTC
251 NCGATGTCGT CATCGAAACT TTTGCCTGCG ACCTGCCCGA AAATGTGCTG
301 CACATCATCC GCCGACACAA GCCGCTTTGG CTGAANTGGG AATATTTGAG
351 CGCGGAGGAN AGCAATGAAA GGCTGCACNT GATGCCTTCG CCGCAGGAGA
401 GTGTTCAAAA ATANTTTTGG TTTATGGGTT TCAGCGAANN NAGCGGCGGA
451 CTGATACGCG AACGCGATTA CTGCGAAGCC GTCCGTTTCG ATAGCGGAGC
501 TTTGCGCAAG AGGCTGATGC TTCCCGAAAA AAACGNCCCC GAATGGCTGC
551 CTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
601 CAGGCAGGCA GTCCGTTGAC ACTTTTGCTG GCNNGGGGCG ANATTATCGA
651 CAGCCTCAAA CAAAACGGCG TTATTCCGCA AGATGCCCTG CAAAACGACG
701 GCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG
751 CCGCAACAGG ACTTCGACAA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT

```

-153-

5  
 801 CCGCGGCGAA GACAGTTTCG TCGCGGCCCA GCTTGC GGCG AAACCTTCT  
 851 TTTGGCACAT CTACCCGCAA GATGAGAATG TCCATCTCGA CAACTCCAC  
 901 GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGCATCGGC  
 951 ACACCGCCGC CTTTCAGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA  
 1001 CACAACGCCT CGAATGTTGG CAAATCCTGC AACACATCA AACCGCTGG  
 1051 CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTTGGGC AGCCTTCCGC  
 1101 ATCCGAAAAA CTCGCCGCTT TTGTTTCAAA GCATCAAAAA ATACGCTAG

This encodes a protein having amino acid sequence <SEQ ID 192>:

10  
 1 MNTPPFSAGX FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR  
 51 ALCPDLPDVX CVHQDIHVRT WHSDAADIDT APVXDVIET FACDLPENVL  
 101 HIIRRHKPLW LXWEYLSAEX SNERLHXMPs PQESVXKXFW FMGFSEXS GG  
 151 LIRERDYCEA VREFD SGALRK RLMLPEKNXP EWLLFGYRSD VWAKWLEMWR  
 201 QAGSPLTLLL AGAXIIDSLK QNGVIPQDAL QNDGDVFQTA SVRLVKIPFV  
 251 PQQDFDKLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPO DENVHLDKLH  
 15  
 301 AFWDKAHGFY TPETASAHRR LSDDLNGGEA LSATQRLCW QILQQHQNGW  
 351 RQGAEDWSRY LFGQPSASEK LAAFVSKHQK IR\*

ORF32a and ORF32-1 show 93.2% identity in 382 aa overlap:

		10	20	30	40	50	60
20	orf32-1.pep	MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX					
	orf32a	MNTPPFSAGXFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX					
		10	20	30	40	50	60
25	orf32-1.pep	CVHQDIHVRTWHSDAADIDTAPVDPVVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEE					
	orf32a	CVHQDIHVRTWHSDAADIDTAPVXDVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX					
		70	80	90	100	110	120
30	orf32-1.pep	SNERLHLMPS PQEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRERLMLPEKNAS					
	orf32a	SNERLHXMPSPQESVXKXFWFMGFSEKSGGLIRERDYCEAVRFD SGALRKRLMLPEKNXP					
		130	140	150	160	170	180
35	orf32-1.pep	EWLLFGYRSDVWAKWLEMWRQAGSPMTLLLAGTQIIDS LKQSGVIPQDALQNDGDVFQTA					
	orf32a	EWLLFGYRSDVWAKWLEMWRQAGSPLTLLLAGAXIIDSLKQNGVIPQDALQNDGDVFQTA					
40		190	200	210	220	230	240
45	orf32-1.pep	SVRLVKIPFVPQQDFDQLHLADCAVIRGEDSFVRAQLAGKPFFWHIYPODENVHLDKLH					
	orf32a	SVRLVKIPFVPQQDFDKLLHLADCAVIRGEDSFVRAQLAGKPFFWHIYPODENVHLDKLH					
		250	260	270	280	290	300
50	orf32-1.pep	AFWDKAHGFYTPETVSAHRRLSDDLNGGEALSATQRLCWQTLQQHQNGWRQGAEDWSRY					
	orf32a	AFWDKAHGFYTPETASAHRRLSDDLNGGEALSATQRLCWQILQQHQNGWRQGAEDWSRY					
		310	320	330	340	350	360
55	orf32-1.pep	LFGQPSAPEKLAAFVSKHQKIRX					
	orf32a	LFGQPSASEKLAAFVSKHQKIRX					
		370	380				

# 60 Homology with a predicted ORF from *N.gonorrhoeae*

ORF32 shows 95.1% identity over a 82aa overlap with a predicted ORF (ORF32.ng) from *N. gonorrhoeae*:

orf32.pep	MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLP	57

```

orf32ng      MVMNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLP      60
orf32.pep    DVPCVHQDIHVRTWHSDAADIDTA                                          81
              ||| ||||||||||||||||||||
5 orf32ng      DVPFVHQDIHVRTWHSDAADIDTAPVPDAVIETFACDLPENVLNIIRRHKPLWLNWEYLS 120

```

An ORF32ng nucleotide sequence <SEQ ID 193> was predicted to encode a protein having amino acid sequence <SEQ ID 194>:

```

10 1 MVMNTYAFPV CWIFCKVIDN FGDIGVSWRL ARVLHRELGW QVHLWTDVVS
    51 ALRALCPDLP DVPFVHQDIH VRTWHSDAAD IDTAPVPDAV IETFACDLP
    101 NVLNIIRRHK PLWLNWEYLS AESNERLHL MPSPQEGVQK YFWFMGFSEK
    151 SGGLIRERDY REAVRFDTEA LRRRLVLPEK NAPEWLLFGY RGDVWAKWLD
    201 MWQQAGSLMT LLLAGAQIID SLKQSGVIPQ NALQNEGGVF QTASVRLVKI
    251 PFVPPQDFDK LLHLADCAVI RGEDSFVRTQ LAGKPFWHI YPDENVHLD
    301 KLHAFWDKAY GFYTPETASV HRLSDDLNG GEALSATQRL ECGVL*

```

15 Further sequencing revealed the following DNA sequence <SEQ ID 195>:

```

20 1 ATGAATACAT ACGCTTTTCC TGTCTGTTGG ATTTTTTGCA AGGTCATCGA
    51 CAATTCGGC GACATCGGCG TTTCTGTCGG GCTCGCCCGT GTTTTCACCC
    101 GCGAAGCTCG TTGGCAGGTG CATTGTGTGA CGGACGACGT GTCCGCCTTG
    151 CGCGCGCTTT GTCCCGATTT GCCCGATGTT CCCTTCGTTC ATCAGGATAT
    201 TCATGTCCGC ACTTGGCATT CCGATGCGGC AGACATTGAT ACCGCGCCCG
    251 TTCCCGATGC CGTTATCGAA ACTTTTGCTT GCGACCTGCC CGAAAATGTG
    301 CTGAACATCA TCCGCCGACA CAAACCGCTT TGGCTGAATT GGAATATTT
    351 GAGCGCGGAG GAAAGCAATG AAAGGCTGCA CCGTATGCCT TCGCCGACAG
    401 AGGGCGTTCA AAAATATTTT TGGTTTATGG GTTTCAGCGA AAAAAGCGGC
    451 GGGTTGATAC GCGAACGCGA TTACCGCGAA GCCGTCCGTT TCGATACCGA
    501 AGCCCTGCGC CGGCGGCTGG TGCTGCCCGA AAAAAACGCC CCCGAATGGC
    551 TGCTTTTCGG CTATCGGGGC GATGTTTGGG CAAAGTGGCT GGACATGTGG
    601 CAACAGGCAG GCAGCCTGAT GACCCTACTG CTGGCGGGGG CGCAAATTAT
    651 CGACAGCCTC AAACAAAGCG GCGTTATTCC GCAAACGCC CTGCAAAAtg
    701 aaggcgGTGT CTTTCagacG gcatccgTcC gccttGTCAA AAtcCCGTTC
    751 GTGCCcCAAC AGGAcTTCGA CAAATGCTG CAcctcgcCG ACTGCGCCGT
    801 GATACGCGGC GAAGACAGTT TCGTGCCTAC CCAGCTTGCC GGAAAACCTT
    851 TTTTTTGGCA CATCTACCG CAAGACGAGA ATGTCCATCT CGACAAACTC
    901 CACGCCTTTT GGGATAAGGC ATACGGCTTC TACACGCCCG AAACCGCATC
    951 GGTGCACCGC CTCCTTTTCGG ACGACCTCAA CGGCGGAGAG GCTTTATCCG
    1001 CAACACAACG CCTCGAATGT TGGCAAACCC TGCAACAACA TCAAAACGGC
    1051 TGGCGGCAAG GCGCGGAGGA TTGGAGCCGT TATCTTTTCG GGCAGCCTTC
    1101 CGCATCCGAA AAACCTCGCG CTTTGTGTTT AAAGCATCAA AAAATACGCT
    1151 AG

```

40 This encodes a protein having amino acid sequence <SEQ ID 196; ORF32ng-1>:

```

45 1 MNTYAFPVCW IFCKVIDNEF DIGVSWRLAR VLHRELGWQV HLWTDVVSAL
    51 RALCPDLPDV PFDVHQDIHVR TWHSDAADID TAPVPDAVIE TFACDLPENV
    101 LNIIRRHKPL WLNWEYLSAE ESNERLHLMP SPQEGVQKYF WFMGFSEKSG
    151 GLIRERDYRE AVRFDEALR RRLVLPEKNA PEWLLFGYRG DVWAKWLDWMW
    201 QQAGSLMTLL LAGAIIIDSL KQSGVIPQNA LQNEGGVFQT ASVRLVKIPF
    251 VPQQDFDKLL HLADCAVIRG EDSFVRTQLA GKPFWHIYP QDENVHLDKL
    301 HAFWDKAYGE YTPETASVHR LLSDDLNGGE ALSATQRLC WTLQQHQNG
    351 WRQGAEDWSR YLFGQPSASE KLAAFVSKHQ KIR*

```

ORF32ng-1 and ORF32-1 show 93.5% identity in 383 aa overlap:

```

50 orf32-1.pep      10      20      30      40      50      59
    MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDV
    ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||
orf32ng-1          10      20      30      40      50      60
    MNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDV

55 orf32-1.pep      60      70      80      90      100     110     119
    PCVHQDIHVRTWHSDAADIDTAPVPDVVIETFACDLPENVLHIIRRHKPLWLNWEYLSAE
    | |||||||||||||||||||||:|||||||||||||:|||||||||||||
orf32ng-1          60      70      80      90      100     110     120
    PFDVHQDIHVRTWHSDAADIDTAPVPDAVIETFACDLPENVLNIIRRHKPLWLNWEYLSAE

60                  120     130     140     150     160     170     179

```



-155-

	orf32-1.pep	ESNERLHLMPS PQEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRERLMLPEKNA
	orf32ng-1	ESNERLHLMPS PQEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRRRLVLPEKNA
5		130 140 150 160 170 180
	orf32-1.pep	180 190 200 210 220 230 239
	orf32ng-1	SEWLLFGYRSDVWAKWLEMWQAGSPMTLLLAGTQIIDSLKQSGVIPQDALQNDGDVFQT
10		190 200 210 220 230 240
	orf32-1.pep	240 250 260 270 280 290 299
	orf32ng-1	PEWLLFGYRGDVWAKWLDWQQAGSLMTLLLAGAQIIDSLKQSGVIPQNALQNEGGVFQT
15		250 260 270 280 290 300
	orf32-1.pep	ASVRLVKIPFVPQQDFDQLHLADCAVIRGEDSFVRAQLAGKPFFFWHIYPQDENVHLDKL
	orf32ng-1	ASVRLVKIPFVPQQDFDQLHLADCAVIRGEDSFVRTQLAGKPFFFWHIYPQDENVHLDKL
20		300 310 320 330 340 350 359
	orf32-1.pep	HAFWDKAHGFYTPETVSAHRRLSDDLNGGEALSATORLECWQTLQQHQNGWRQGAEDWSR
	orf32ng-1	HAFWDKAYGFYTPETASVHRRLSDDLNGGEALSATORLECWQTLQQHQNGWRQGAEDWSR
25		360 370 380
	orf32-1.pep	YLFQGPSAPEKLAAFVSKHQKIRX
	orf32ng-1	YLFQGPSASEKLAAFVSKHQKIRX
		370 380

- 30 On this basis, including the RGD sequence in the gonococcal protein, characteristic of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF32-1 (42kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 35 7A shows the results of affinity purification of the His-fusion protein, and Figure 7B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA, giving a positive result. These experiments confirm that ORF32-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 24

- 40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 197>:

1 ..TTGTTCCCTGC GTGTNAAAGT GGGGCGTTTT TTCAGCAGTC CGGCGACGTG  
51 GTTTCGGGNC AAAGACCCCTG TAAATCAGGC GGTGTTGCGG CTGTATNCGG  
101 ACGAGTGGCG GCA.ACTTCG GTACGTTGGA AAATAGNCGC AACGTCGCAC  
151 AGCCTGTGGC TCTGCACGCT GCTCGGAATG CTGGTGTCGG TATTGTTGCT  
45 201 GCTTTTGGTG CGGCAATATA CGTTCAACTG GGAAAGCACG CTGTTGAGCA  
251 ATGCCGCTTC GGTACGCGCG GTGGAATGT TGGCATGGCT GCCGTCGAAA  
301 CTCGGTTTCC CTGTCCCCGA TGCGCGGTGCG GTCATCGAAG GCCGCTCTGAA  
351 CGGCAATATT GCCGATGCGC GGGCTTGGTC GGGGCTGCTG GTCGNCAGTA  
401 TCGCTGCTA NGGCATCCTG CCGCGCCTG..

- 50 This corresponds to the amino acid sequence <SEQ ID 198; ORF33>:

1 ..LFLRVKVGFR FSSPATWFRX KDPVNQAVLR LYXDEWRXTS VRWKIXATSH  
51 SLWLCTLLGM LVSLLLLLV RQYTFNWEST LLSNAASVRA VEMLAWLPSK  
101 LGFPVPDARS VIEGRLNGNI ADARAWSGLL VXSIACXGIL PRL..

Further work revealed the complete nucleotide sequence <SEQ ID 199>:

```

1  ATGTTGAATC CATCCCGAAA ACTGGTTGAG CTGGTCCGTA TTTTGGACGA
51 AGGCGGTTTT ATTTTCAGCG GCGATCCCGT ACAGGCGACG GAGGCTTTGC
101 GCCGCGTGGA CGGCAGTACG GAGGAAAAAA TCATCCGTCG GCGCGAGATG
5  151 ATTGACAGGA ACCGTATGCT GCGGGAGACG TTGGAACGTG TCGGTGCGGG
201 GTCGTTCTGG TTGTGGGTGG TGGCGGCGAC GTTTGCATTT TTTACCGGTT
251 TTTCACTCAC TTATCTTCTA ATGGACAATC AGGGTCTGAA TTTCTTTTGG
301 GTTTTGGCGG GCGTGTGGG CATGAATACG CTGATGCTGG CAGTATGGTT
351 GGCAATGTTG TTCCTGCGTG TGAAAGTGGG GCGTTTTTTC AGCAGTCCGG
10  401 CGACGTGGTT TCGGGGCAAA GACCTGTAA ATCAGGCGGT GTTGCGGCTG
451 TATGCGGACG AGTGGCGGCA ACCTTCGGTA CGTTGAAAA TAGGCGCAAC
501 GTCGCACAGC CTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTCGGTAT
551 TGTTGCTGCT TTTGGTGGG CAATATACGT TCAACTGGGA AAGCAGCTG
601 TTGAGCAATG CCGCTTCGGT ACGCGCGGTG GAAATGTTGG CATGGCTGCC
15  651 GTCGAAATC GGTTCCTCG TCCCGATGC GCGGCGGTC ATCGAAGGCC
701 GTCTGAACGG CAATATTGCC GATGCGCGGG CTTGGTGGG GCTGCTGGT
751 GGCAGTATCG CCTGCTACGG CATCTGCCG CGCCTGCTGG CTTGGGTAGT
801 GTGTAATC CTTTTGAAAA CAAGCGAAA CGGATTGGAT TTGAAAAGC
851 CCTATTATCA GGCGGTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG
20  901 GATACGCGTC GGGAAACCGT GTCCGCCGTT TCACCGAAAA TCATCTTGAA
951 CGATGCGCCG AAATGGGCGG TCATGCTGGA GACCGAGTGG CAGGACGGCG
1001 AATGGTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
1051 ACCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
1101 GGCGCAACTG CTTATCGCGG TGCGCGCCA AACTGTGCCG GACCGCGGCG
25  1151 TGTTGCGGCA GATTGTCCGA CTCTCGGAAG CGGCGCAGGG CGGCGCGGTG
1201 GTGCAGCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
1251 GGAACATTGG CGTAACGCGC TGGCCGAATG CGGCGCGGCG TGGCTTGAGC
1301 CTGACAGGGC GGCGCAGGAA GGGCGTTTGA AAGACCAATA A

```

This corresponds to the amino acid sequence <SEQ ID 200; ORF33-1>:

```

30  1  MLNPSRKLVE LVRILDEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAEM
51  IDRNRMLRET LERVAGSEF LWVVAATFAF FTGFSVTYLL MDNQGLNFFL
101 VLAGVLGMNT LMLAVWLAML FLRVKVRFF SSPATWFRGK DPNVQAVLRL
151 YADEWRQPSV RWKIGATSHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
35  201 LSNAASVRV EMLAWLPSKL GFPVPDARV IEGRNLGNIA DARAWSGLLV
251 GSIACYGILP RLLAWVCKI LLKTSENGLD LEKPYQAVI RRWQNKITDA
301 DTRRETSAV SPKII LDAP KWAVMLETEW QDGEWFEGRL AQEWLDKGVA
351 TNREQVAAL TELKQKPAQL LIGVRAQTVP DRGVLQIVR LSEAAQGGAV
401 VQLLAEQGLS DDLSEKLEHW RNALAECEGAA WLEPDRAAQE GRLKDQ*

```

Computer analysis of this amino acid sequence gave the following results:

#### 40 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF33 shows 90.9% identity over a 143aa overlap with an ORF (ORF33a) from strain A of *N. meningitidis*:

```

45  orf33.pep                                10      20      30
                                         LFLRVKVRFFSSPATWFRXKDPVNQAVLR
                                         |||
orf33a  LMDNQGLNFFLVLAGVXGMNTLMLAVWLAMLFLRVKVRFFSSPATWFRGKDPVNQAVLR
          90      100      110      120      130      140

50  orf33.pep          40      50      60      70      80      90
LYXDEWRXTSVRWKIXATSHSLWLCTLLGMLVSVLLLLLVRQYTFNWESTLLSNAASVRA
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
orf33a  LYADEWRXPSVRWKIGATSHSLWLCTLLGMLVSVLLLLLVRQYTFNWESTLLGDSSSVRL
          150      160      170      180      190      200

55  orf33.pep          100      110      120      130      140
VEMLAWLPSKLGFPVPDARSVIEGRNLGNIA DARAWSGLLVXSIACXGILPRL
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
orf33a  VEMLAWLPAKLGFPVPDARAVIEGRNLGNIA DARAWSGLLVGSIACYGILPRL LAWAVCK
          210      220      230      240      250      260

60  orf33a  ILXXTSENGLDLEKXXXXXXIRRWQNKITDADTRRETSAVSPKIVLNDAPKWAVMLETE
          270      280      290      300      310      320

```

The complete length ORF33a nucleotide sequence <SEQ ID 201> is:

```

1  ATGTGAATC CATCCCGAAA ACTGTTGAG CTGGTCCGTA TTTTGAAGA
51 AGGCGGCTTT ATTTTCAGCG GCGATCCCGT GCAGGCGACG GAGGCTTTGC
101 GCCGCGTGGA CGGCAGTACG GAGGAAAAAA TCATCCGTCG GGCGAAGATG
5   151 ATCGACAGGA ACCGTATGCT GCGGGAGACG TTGGAACGTG TGCGTGCGGG
201 GTCGTTCTGG TTGTGGGTGG CGGCGGCGAC GTTTGCGTTT NTTACCGNNT
251 TTTAGATTAC TTATCTTCTA ATGGACAATC AGGGTCTGAA TTTCTTTTGT
301 TTTTGGCGGG GCGTGNTGGG CATGAATACG CTGATGCTGG CAGTATGGTT
351 GGCAATGTTG TTCCTGCGCG TGAAAGTGGG GCGTTTTTTC AGCAGTCCGG
10  401 CGACGTGGTT TCGGGGCAAA GACCCTGTCA ATCAGGCGGT GTTGCGGCTG
451 TATGCGGACG AGTGGCGGCN ACCTTCGGTA CGTTGAAAAA TAGGCGCAAC
501 TGTGCACAGC CTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTCGGTAT
551 GTTGCTGCTT TTGGTGCGG CAATATACGT TCAACTGGGA AAGCACGCTG
601 TTGGGCGATT CGTCTTCGGT ACGGCTGGTG GAAATGTTGG CATGGCTGCC
15  651 TGCGAAACTG GGTTTTCCCG TGCCTGATGC GCGGGCGGTC ATCGAAGGTC
701 GTCTGAACGG CAATATTGCC GATGCGCGGG CTTGGTTCGGG GCTGCTGGTC
751 GGCAGTATCG CCTGCTACGG CATCCTGCCG CGCCTCTTGG CTTGGGCGGT
801 ATGCAAAATC CTTNTGNAAA CAAGCGAAAA CGGCTTGGAT TTGAAAAGC
851 NNNNNNTCN NNCGNTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG
20  901 GATACGCGTC GGGAAACCGT GTCCGCCGTT TCGCCGAAAA TCGTCTTGAA
951 CGATGCGCCG AAATGGGCGG TCATGCTGGA GACCGAATGG CAGGACGGCG
1001 AATGTTTCGA GGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
1051 GCCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
1101 GGCGCAACTG CTTATCGGCG TCGCGCCCA AACTGTGCC GACCGCGGCG
25  1151 TGTGCGGCA GATCGTCCGA CTTTCGGAAG CGGCGCAGG CGGCGCGGTG
1201 GTGCANCTTT TGGCGGAACA GGGGCTTCA GACGACCTT CCGAAAAGCT
1251 GGAACATTGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTGGAC
1301 CCGACAGAGC GGCGCAGGAA GGCCGTCTGA AAACCAACGA CCGCACTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 202>:

```

30  1  MLNPSRKLV LVRILEEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAKM
51  IDNRNMLRET LERVAGSEFW LWVAAATFAE XTXFSVTYLL MDNQGLNFFL
101 VLAGVXGMNT LMLAVWLAML FLRVKVGREF SSPATWFRGK DPNVQAVLRL
151 YADEWRXPSV RWKIGATSHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
201 LGDSSSVRLV EMLAWLPACL GFPVPDARAV IEGRNLGNIA DARAWSGLLV
35  251 GSIACYGILP RLLAWAVCKI LXXTSENGLD LEKXXXXXXI RRWQNKITDA
301 DTRRETVS AV SPKIVLNDAP KWAVMLETEW QDGEWFEGRL AQEWLDKVA
351 ANREQVALE TELKQKPAQL LIGVRAQTVP DRGVLRQIVR LSEAAQGGAV
401 VXLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRAAQE GRLKTNDR*

```

ORF33a and ORF33-1 show 94.1% identity in 444 aa overlap:

```

40      10      20      30      40      50      60
orf33a.pep  MLNPSRKLV LVRILEEGGF IFSGDPVQATEALRRVDGST EEKIIRRAKMIDNRNMLRET
          |||
orf33-1     MLNPSRKLV LVRILEEGGF IFSGDPVQATEALRRVDGST EEKIIRRAEMIDNRNMLRET
          |||

45      10      20      30      40      50      60
orf33a.pep  LERVAGSEFW LWVAAATFAE XTXFSVTYLLMDNQGLNFFL VLAGVXGMNT LMLAVWLAML
          |||
orf33-1     LERVAGSEFW LWVAAATFAE XTXFSVTYLLMDNQGLNFFL VLAGVXGMNT LMLAVWLAML
          |||

50      70      80      90      100     110     120
orf33a.pep  LERVAGSEFW LWVAAATFAE XTXFSVTYLLMDNQGLNFFL VLAGVXGMNT LMLAVWLAML
          |||
orf33-1     LERVAGSEFW LWVAAATFAE XTXFSVTYLLMDNQGLNFFL VLAGVXGMNT LMLAVWLAML
          |||

55      130     140     150     160     170     180
orf33a.pep  FLRVKVGREF SSPATWFRGK DPNVQAVLRL YADEWRXPSV RWKIGATSHS LWLCTLLGML
          |||
orf33-1     FLRVKVGREF SSPATWFRGK DPNVQAVLRL YADEWRXPSV RWKIGATSHS LWLCTLLGML
          |||

60      190     200     210     220     230     240
orf33a.pep  VSVLLLLLVR QYTFNWESTL LGDSSSVRLV EMLAWLPACL GFPVPDARAV IEGRNLGNIA
          |||
orf33-1     VSVLLLLLVR QYTFNWESTL LLSNAASVRAV EMLAWLPACL GFPVPDARAV IEGRNLGNIA
          |||

65      250     260     270     280     290     300
orf33a.pep  DARAWSGLLV GSIACYGILP RLLAWAVCKI LXXTSENGLD LEKXXXXXXI RRWQNKITDA
          |||

```

-158-

	orf33-1	DARAWSGLLVGS	IACYGILPRLLAWVVCKILLKTS	ENGLDLEKPYQAVIRRWNKITDA	
		250	260	270	280
		290	300		
5	orf33a.pep	310	320	330	340
		350	360		
	orf33-1	DTRRET	VS	SAVSPKIVLNDAPKWAVMLETEWQDGEWFEGRLAQEWLDKGV	AANREQVAAL
		310	320	330	340
		350	360		
10	orf33a.pep	370	380	390	400
		410	420		
	orf33-1	TELKQKPAQLLIGVRAQTVPDRGVLRQIVRLSEAAQGGAVVXLLAEQGLSDDLSEKLEHW			
		370	380	390	400
		410	420		
15	orf33a.pep	430	440	450	
		RNALTECGAAWLEPDRAAQEGR	LKTNDRTX		
	orf33-1	RNALAECGAAWLEPDRAAQEGR	LKDQX		
20		430	440		

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF33 shows 91.6% identity over a 143aa overlap with a predicted ORF (ORF33.ng) from *N. gonorrhoeae*:

25	orf33.pep	LFLRVKVGRRFFSSPATWFRXKDPVNQAVLR	30
	orf33ng	LMDNQGLNFFLVLAGVLGMNTLMLAVWLATLFLRVKVGRRFFSSPATWFRGKGPVNQAVLR	100
30	orf33.pep	LYXDEWRXTSVRWKIXATSHSLWLCTLLGMLVSVLLLLLVRQYTFNWESTLLSNAASVRA	90
	orf33ng	LYADQWRQPSVRWKIGATAHSLWLCTLLGMLVSVLLLLLVRQYTFNWESTLLSNAASVRA	160
	orf33.pep	VEMLAWLPSKLGFPVPDARSVIEGR	143
35	orf33ng	VEMLAWLPSKLGFPVPDARAVIEGR	220

An ORF33ng nucleotide sequence <SEQ ID 203> was predicted to encode a protein having amino acid sequence <SEQ ID 204>:

	1	MIDRDRMLRD	TLERVAGSF	WLWVVASMM	FTAGFSGTYL	LMDNQGLNFF
	51	LVLAGVLGMN	TLMLAVWLAT	LFLRVKVGRR	FSSPATWFRG	KGPVNQAVLR
40	101	LYADQWRQPS	VRWKIGATAH	SLWLCTLLGM	LVSLLLLLV	ROYTFNWEST
	151	LLSNAASVRA	VEMLAWLPSK	LGFPVPDARA	VIEGR	LNGNI ADARAWSGLL
	201	VGSIVCYGIL	PRLLAWVVCK	ILLKTS	ENGL DLEKTYQAV	IRRWNKITD
	251	ADTRRET	VSA VSPKIVLND	APKWALMLETE	WQDQWFEGR	LAQEWLDKGV
	301	AANREQVAAL	ETELKQKPAQ	LLIGVRAQTV	PDRGVLRQIV	RLSEAAQGGA
45	351	VVQLLAEQGL	SDDLSEKLEH	WRNALTECGA	AWLEPDRAVQ	EGR

Further sequence analysis revealed the following DNA sequence <SEQ ID 205>:

	1	ATGTTGaatC	CATCCGaaAA	ACTGgttgag	ctGgTCCgtA	Ttttgaataa
	51	agggggtTTT	attttcagcg	gcgatcctgt	gcaggcgacg	gaggctttgc
50	101	gccgcgtgga	cggcAGTACG	GAGgAaaaaa	tcttccgtcg	GGCGGAGAtg
	151	atcgACAGGg	accgtatggt	gcgggACaCg	TtggaacGTG	TGCGTGCGgg
	201	gtcgtTctgG	TTATGGGTGG	TggtggCaTc	gATGATGTtt	aCCGCCGGAT
	251	TTTCAGgcac	ttatCtctCTG	ATGGACaatC	AGGGGcTGA	TtTCTTTTTA
	301	GTTTTggcgG	GAGTGTtggG	CATGaatacG	ctgATGCTGG	CAGTATGGtt
	351	gGCAACGTTG	TTCCTGCGCG	TGAAAGTGGG	ACGGTTTTTT	AGCAGTCCGG
55	401	CGACGTGGT	TCGGGGCAAA	GGCCTGTAA	ATCAGGCGGT	GTTGCGGCTG
	451	TATGCGGACC	AGTGGCGGCA	ACCTTCGGTA	CGATGGAAAA	TAGGCGCAAC
	501	GGCGCACAGC	TTGTGGCTCT	GCACGCTGCT	CGGAATGCTG	GTGTGCGGTAT
	551	TGCTGCTGCT	TTTGGTGCGG	CAATATACGT	TCAACTGGGA	AAGCACGCTG
	601	TTGAGCAATG	CCGCTTCGGT	ACGCGCGGTG	GAAATGTTGG	CATGGCTGCC
60	651	GTCGAAACTC	GGTTTCCTG	TCCCGATGC	GCGGGCGGTC	ATCGAAGGTC
	701	GTCTGAACGG	CAATATTGCC	GATGCGCGGG	CTTGGTCGGG	GCTGCTGGTC
	751	GGCAGTATCG	TCTGCTACGG	CATCCTGCCG	CGCCTCTTGG	CTTGGGTAGT

-159-

801 GTGTAAAATC CTTTGTAAAA CAAGCGAAAA CGGATTgGAT TTGGAAAAAA  
 851 CCTATTATCA GGCGGTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG  
 901 GATACGCGTC GGGAAACCGT GTCCGCGGTT TCGCcgAAAA TCGTCTTGAA  
 951 CGATGCGCCG AAATGGGCGC TCATGCTGGA GACCGAGTGG CAGGACGGCC  
 1001 AATGGTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC  
 1051 GCCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC  
 1101 GGCGCAACTG CTTATCGGCG TACGCGCCCA AACTGTGCCG GACCGGGGCG  
 1151 TGCTGCGGCA GATTGTGCGG CTTTCGGAAG CGGCGCAGGG CGGCGCGGTG  
 1201 GTGCAGCTTT TGGCGGAACA GGGGCTTCA GACGACCTTT CGGAAAAGCT  
 1251 GGAACATTGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTTGAGC  
 1301 CTGACAGGGT GGCGCAGGAA GGCCGTTTGA AAGACCAATA A

This encodes a protein having amino acid sequence <SEQ ID 206; ORF33ng-1>:

1 MLNPSRKLVE LVRILNKGGF IFSGDPVQAT EALRRVDGST EEKIFRAEM  
 51 IDRDRLRDT LERVAGSEFW LWVVASMMF TAGFSGTYLL MDNQGLNFFL  
 101 VLAGVLGMNT LMLAVWLATL FLRVKVGRRF SSPATWFRGK GPVNQAVLRL  
 151 YADQWRQPSV RWKIGATAHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL  
 201 LSNAASVRV EMLAWLPSKL GFPVPDARV IEGRNLGNIA DARAWSGLLV  
 251 GSIVCYGILP RLLAWVVCKI LLKTSENGLD LEKTYQAVI RRWQNKITDA  
 301 DTRRETSAV SPKIVLNDAP KWALMLETW QDQWFEGRL AQEWLDKGVA  
 351 ANREQVALE TELKQKPAQL LIGVRAQTV DRGVLRIQVR LSEAAQGGAV  
 401 VQLLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRVAQE GRLKDQ\*

ORF33ng-1 and ORF33-1 show 94.6% identity in 446 aa overlap:

		10	20	30	40	50	60
25	orf33-1.pep	MLNPSRKLVELVRILDEGGFIFSGDPVQATEALRRVDGST	EEKIIIRRAEMIDRNRLRET				
	orf33ng-1	MLNPSRKLVELVRILNKGGFIFSGDPVQATEALRRVDGST	EEKIFRAEMIDRDRLRDT				
		10	20	30	40	50	60
30	orf33-1.pep	LERVRAGSFWLWVVAATFAFFTGFSVTYLLMDNQGLNFFLVL	LAGVLGMNTLMLAVWLAML				
	orf33ng-1	LERVRAGSFWLWVVASMMFTAGFSGTYLLMDNQGLNFFLVL	LAGVLGMNTLMLAVWLATL				
		70	80	90	100	110	120
35	orf33-1.pep	FLRVKVGRRFFSSPATWFRGKDPVNQAVLRLYADEWRQPSV	RWKIGATSHSLWLCTLLGML				
	orf33ng-1	FLRVKVGRRFFSSPATWFRGKGPVNQAVLRLYADQWRQPSV	RWKIGATAHSLWLCTLLGML				
		130	140	150	160	170	180
40	orf33-1.pep	VSVLLLLLVRQYTFNWESTLLSNAASVRVAVEMLAWLPSKL	GFPVPDARAVIEGRNLGNIA				
	orf33ng-1	VSVLLLLLVRQYTFNWESTLLSNAASVRVAVEMLAWLPSKL	GFPVPDARAVIEGRNLGNIA				
45		190	200	210	220	230	240
50	orf33-1.pep	DARAWSGLLVGSIACYGILPRLAWVVCKILLKTSENGLDLEK	PYYQAVIRRWQNKITDA				
	orf33ng-1	DARAWSGLLVGSIACYGILPRLAWVVCKILLKTSENGLDLEK	TYQAVIRRWQNKITDA				
		250	260	270	280	290	300
55	orf33-1.pep	DTRRETSAVSPKIIINDAPKWAVMLETWQDGEWFEGRLAQE	WLDKGVAATNREQVALE				
	orf33ng-1	DTRRETSAVSPKIVLNDAPKWALMLETWQDGEWFEGRLAQE	WLDKGVAANREQVALE				
		310	320	330	340	350	360
60	orf33-1.pep	TELKQKPAQLLIGVRAQTVPDGRVLRQIVRLSEAAQGGAVV	QLLAEQGLSDDLSEKLEHW				
	orf33ng-1	TELKQKPAQLLIGVRAQTVPDGRVLRQIVRLSEAAQGGAVV	QLLAEQGLSDDLSEKLEHW				
		370	380	390	400	410	420
65	orf33-1.pep	RNALAECEGAAWLEPDRAAQEGRLKDQX					
		430	440				

orf33ng-1                      RNALTECGAAWLEPDRVAQEGRLKDQX  
    430                      440

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 25

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 207>:

```

10      1  ..CAGAAGAGTT TGTCGAGAAT TTCTTTATGG GGTTTGGGCG GCGTGTTTTT
      51  CGGGGTGTCC GGTCTGGTAT GGTTTCTTTT GGGCGTTTCT TT.GAGTGCG
     101  CCTGTTTTTC GGGTGTCTT TTTCGGGGT CGGGACGGG GACGTTTGTG
     151  GGCAGTACGG GGGTTCTTT GAGTGTGTTT TCAGCTTG TG .GGCGT
     201  CGTCCGGCTG CCTGTGCGT TGAGCTGTG CGGCAGGTTG CG..GTTTGA
     251  CCCGGTTTTT CTGGGTGCG GCAGGGGACG TCATTCTCCT GCCGCTTTCG
     15  301  TCTGTGCCGT CCGGCTGTG GGGTTCGGAT GAGGCGGCGT GGTGGTGTTT
     351  GGGTTGGGCG GCATCTTGTT CCGACTACGC CGTTTGGCAG CCAGAATTCTG
     401  GTTTCGCGGG GGCTGTCGGT GTGTGCGGT TCGGCTTGAA GGGTTTTGTC
     451  GTCC..
  
```

This corresponds to the amino acid sequence <SEQ ID 208; ORF34>:

```

20      1  ..QKSLSRISLW GLGGVFFGVVS GLVWFSLGVS XECACFSGVS FRGSGRGTFV
     51  GSTGVSLSVF SACVXGVVRL PVGLSCVGRL XXLTRFFLGA AGDVILLPLS
     101  SVPSGCAGSD EAAWWC SGWA ASCPTTPFGS QNSVSRGLSV CCGSA*RVLS
     151  S..
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 209>:

```

25      1  ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGCKGGTG TGCCTGCCGT
     51  GCCGGGTCAG AATAGGTTGT CCAGAAATTC TTTATGGGGT TTGGGCGGCG
     101  TGTTTTTCGG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTTTG
     151  GGCTGCGCCT GTTTTTTCGG TGTTTCTTTT CGGGGTTCGG GACGGGGGAC
     201  GTTTGTGGGC AGTACGGGGG TTTCTTTGAG TGTGTTTTCA GCTTGTGTTC
     30  251  CGGCGTCGTC CGGCTGCCG TCGGTTGAG CTGTGTCGGC AGGTGCGGT
     301  TTGACCCGGT TTTTCTTGGG TGCGGCAGGG GACGGCAGTC CGTGCCGCT
     351  TTCGCTGTG CCGTCCGGCT GTGCGGGTTC GGATGAGGCG GCGTGGTGGT
     401  GTTCGGGTTG GCGGCATCT TGTCCGACTA CGCCGTTTGG CAGCCAGAAT
     451  TCGGTTTCGC GGGGGCTGTC GGTGTGTTGC GGTTGCGCTT GAAGGGTTTT
     35  501  GTCGCCGTTT GGGTTGAATG TGCTGACGAT GCCTATTGCC AATGCGCCGA
     551  TGGCGGCGAT ACAGATGAGC AATACGGCGC GTATCAGGAG TTTGGGGGTC
     601  AGCCTGAAGG GTTTGTTCCG TTTTTTTGCC ATTTTGATTG TGCTTTTGGG
     651  GTGTGCGGCA ATGCCGCTG AAGGCGGTTT AGACGGCATT GCCGATTCAG
     701  CGTTGGACGT AGTTTTGGTA GAGGGTGATG ACTTTTGTGA CGCCGACGGT
     40  751  GGTGCTGACT TTTTGGGTAA TCTGCGCTG TTCTTCGGGG GTGAGGATGC
     801  CCATAACGTA GGTTACGTTG CCGTAGGTAA CGATTTTGAC GCGCGCCTGT
     851  GTGGCGGGGC TGATGCCCAA CAGCGTGGCG CGGACTTTGG ATGTGTTCCA
     901  AGTGTGCGCG GCGATGTCGC CGGCAGTGCG CGGCAGGGAG GCGACGGTAA
     951  TATAGTTGTA CACGCCTTCG GCGGCCTGTT CGGAACGTGC AATCTGACCG
     45  1001  ACGAACTGTT TTTGCGCTTC GGTGGCGACT TGTCCGAGCA GCAGCAGGTG
     1051  GCGGTTGTAG CCGACGACGG AGATTTGGGG CGTGAGCCT TTGGTTTGGT
     1101  TGTTTTGGCG CAGATAGGAA CGGGCGGTGG TTTGATACG CAACGCCATA
     1151  ACGTTGTGCT CGGTTTGGCG GCCGGTGGTT CGGCGGTGCA CGGCGGATTT
     1201  CGCGCCGACG GCGGCGCTTC CGATTACTGC GCTGACGCAG CCGCTAAGGG
     50  1251  CAAGGCTGAA AATGGCGGCA ATCAGGCTGC GGACGGTGTG CGGTTTGGGT
     1301  TTCATCGGGT GCTTCCTTTC TTGGGCGTTT CAGACGGCAT TGCTTTGCGC
     1351  CATGCCGTCT GA
  
```

This corresponds to the amino acid sequence <SEQ ID 210; ORF34-1>:

```

55      1  MPMFIMLPW IAGVPAVPGQ NRLSRISLWG LGGVFFGVSG LVWFSLGVS
     51  GCACFSGVSF RSGRGTFVG STGVSLSVFS ACVPASSGCL SV*AVSAGC
     101  LTRFFLGAAG DGSPLPLSSV PSGCAGSDEA AWWCSGWAAS CPTTPFGSQN
     151  SVSRGLSVCC GSA*RVLSPF GLNVLTMPIA NAPMAAIQMS NTARISLGV
  
```

Computer analysis of this amino acid sequence gave the following results:

**Homology with a predicted ORF from *N.meningitidis* (strain A)**

ORF34 shows 73.3% identity over a 161aa overlap with an ORF (ORF34a) from strain A of *N.*

10 *meningitidis:*

```

10      15      20      25      30
orf34.pep          QKSLSRISLWGLGGVFFGVSGVLWFSLGVSXE-----CAC
|| ||| ||||||| ||||||| ||||||| |||
orf34a      MMXPXIMLPWIAGVPAPVGQKRLSRXSLWGLGGXFFGVSGLVWFSLGVSXSLSGV SXGCAC
              10       20       30       40       50       60

40      50      60      70      80      90
orf34.pep      FSGVSFRGSGRGTFVGSTGVSLSVFSACVXGVVRLPVGLSCVGR LXX----LTRFFLGA
||||| ||||||| ||||||| ||| :| |
orf34a      FSGVSFRGSGRGTFVGSTGVSLSVFSACA-----PASSGCLSVXAVSAGCGLTRFXFGA
              70       80       90      100      110

100     110     120     130     140     150
orf34.pep      AGDVILLPLSSVPSGCAGSDEAAWC SGWAASCTPTPFGSQNSVSRGLSVCCGSAXRVLS
||| ||||||| :|| | ||||||| ||||||| ||| : |||
orf34a      AGDGSPLPLSSVPSGCAGADEEAXXC SGWAASCTPTPFGSQNSVSRGLSVCCGSVWRVLS
              120     130     140     150     160     170

30      orf34.pep      S
orf34a      PFGXNVLTMPIANAPMAVIQMSNTARIRSLGVSLKGLF XFFAILIVLLGCRAMPSEGGSD
              180     190     200     210     220     230

```

The complete length ORF34a nucleotide sequence <SEQ ID 211> is:

35	1	ATGATGATNC	CGTTNATAAT	GCTTCCCTGG	ATTGCGGGTG	TGCTGCGCGT
	51	GCCGGGTGAG	AAGAGGTTGT	CGAGAANTTC	TTTATGGGGT	TTAGGCGGCN
	101	TGTTTTTCGG	GGTGTCCGGT	TTGGTATGGT	TTTCTTTGGG	CGTTTCTNTT
	151	TCTTTGGGTG	TTTCTNTGGG	CTGTGCCTGT	TTTTCGGGTG	TTTCTTTTCG
40	201	GGGTTCGGGA	CGGGGGACGT	TTGTGGGCGT	TACNNGGGTT	TCTTTGAGTG
	251	TGTTTTCCAG	TTGTGCTCCG	GCGTCGTCCG	GCTGCGCTGC	GGCTTTCAGT
	301	GTGTCCGGAG	GTTGCGGTTT	GACCCGGNTT	TTCTTNGGTG	CGGCAGGGGA
	351	CGGCAGTCCG	CTGCCGCTTT	CGTCTGTGCC	GTCCGGCTGT	GCGGGTGCAG
45	401	ATGAGGAGGC	GTNGTNGTGT	TCGGGTTGGG	CGGCATCTTG	TCCGACTACG
	451	CCGTTTGGCA	GCCAGAATTC	GGTTTCGGCG	GGGCTGTCCG	TGTGTTGCGG
	501	TTCCGTTNTG	AGGGTTTTGT	CNCCGTTCGG	TGNAAGATGT	CTGACGATGC
	551	CTATTGCCAA	TGCGCCGATG	GCGGTGATAC	AGATGAGCAA	TACGGCGCGT
50	601	ATCAGGAGTT	TGGGGGTGAG	CCTGAAGGGT	TTGTTTCNGT	TTTTTGCCAT
	651	TTTGATTGTG	CTTTTGGGGT	GTCGGGCAAT	GCCGTCTGAA	GCGGGTTTCA
	701	ACGGCATTGC	CGAGTCAGCG	TTGGACGTAG	TTTNGGTAAG	GGGTGATGAC
	751	TTTTTGTAGC	CCGACGGTGG	TGCTGACTTT	TTGGGTAATG	TGCGCTGTTT
55	801	CTTCGGGGTG	GAGGATGCCG	ATAACGTAGG	TTACGTTGCC	GTAGGTAACG
	851	ATTTTGACGC	GCGCCTGTGT	GGCGGGGCTG	ATGCCCAACA	GCGTGGCGCG
	901	GACTTTGGAT	GTGTTCCAAG	TGTCGCGCGC	GATGTCGCGG	GCACTGCGCG
	951	GCAGGGAGGC	GACGGTAATG	TANTTGTACA	CGCCTTCGGC	GGCCTGTTTC
60	1001	GAACGTGCAA	TCTGACCGAC	GAACGTGTTT	TGCGCTTCGG	TGGCGACTTG
	1051	TCCGAGTACA	AGCAGGTGGC	GGTTGTAGCC	GACACCGGAG	ATTTTGGGGCG
	1101	TCTANCTTTT	GGTTTGGTTG	TTTTGGCGCA	GATAGGAGCG	GGCGGTGGTT
	1151	TCGATACGCA	GCGCCATTAC	GTTGTGCTCG	GTTNGCGCGC	CGGTGGTTTC
65	1201	GCGGTGACGC	GCGGATTTTC	CGCCGACCGC	CGCGCCGCGG	ACGACTGCGC
	1251	TGACGCAGCC	GCCGAGGGCA	AGGCTGAGGA	CGGCGGCAGT	CAGGGTGCAG
	1301	ACGGGTGTCG	GTTTGGGTTT	CATCGGGTGC	TTCTTTTCTT	GGGCGTTTCA
	1351	GACGGCATTG	CTTTGCGCCA	TGCCGTCTGA		

This encodes a protein having amino acid sequence <SEQ ID 212>:

```

      1  MMXPXIMLPW IAGVPAVPGQ KRLSRXSLWG LGGXFFGVSG LVWFSLVGSX
     51  SLGVXGXCAC FSGVSFRGSG RGTfVGSTGV SLSVFSACAP ASSGCLSVXA
    101  VSAGCGLTRX FXGAAGDGSP LPLSSVPSGC AGADEEAXXC SGWAASCPTT
    151  PFGSQNSVSR GLSVCCGSVW RVLSPFGXNV LTMPIANAPM AVIQMSNTAR
    201  IRSLGVSXKG LFXFFAILIV LLGCRAMPSE GSGDGAESA LDVVXVEGDD
    251  FLYADGGADF LGNLRLLFFGG EDAHNVGYVA VGNDFDARLC GGADAQQRGA
    301  DFGCVPSVAG DVAGSARQGG DGNVXVHAFG GLFGTCNLTD ELFLAFGGDL
    351  SEQQVAVVA DNGDLGRVXF GLVLAQIGA GGGFDTQRHY VVGXRAGGS
    401  AVDGGFRADR RAADDCADAA AEGKAEDGGS QGADGVRFGF HRVLPFLGVS
    451  DGIALRHAV*
  
```

ORF34a and ORF34-1 show 91.3% identity in 459 aa overlap:

```

      10      20      30      40      50      60
    15  orf34a.pep  MMXPXIMLPWIAGVPAVPGQKRLSRXSLWGLGGXFFGVSGLVWFSLVGSXSLGVXGXCAC
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    orf34-1  MMMPFIMLPWIAGVPAVPGQNRLSRISLWGLGGVFFGVSGLVWFSLVGSL-----GCAC
      10      20      30      40      50

      70      80      90     100     110     120
    20  orf34a.pep  FSGVSFRGSGRGTfVGSTGVSLSVFSACAPASSGCLSVXAVSAGCGLTRXFXGAAGDGSP
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    orf34-1  FSGVSFRGSGRGTfVGSTGVSLSVFSACVPASSGCLSVXAVSAGCGLTRFFLGAAGDGSP
      60      70      80      90     100     110

      130     140     150     160     170     180
    25  orf34a.pep  LPLSSVPSGCAGADEEAXXC SGWAASCPTT PFGSQNSVSRGLSVCCGSVWRVLSVSPFGXNV
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    orf34-1  LPLSSVPSGCAGSDEAAWCSGWAASCPTT PFGSQNSVSRGLSVCCGSAXRVLSVSPFGLNV
      120     130     140     150     160     170

      190     200     210     220     230     240
    30  orf34a.pep  LTMPIANAPMAVIQMSNTARIRSLGVSLKGLFXFFAILIVLLGCRAMPSEGGSDGAESA
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    orf34-1  LTMPIANAPMAAIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGAESA
      180     190     200     210     220     230

      250     260     270     280     290     300
    35  orf34a.pep  LDVVXVEGDDFLYADGGADFLGNLRLLFFGGEDAHNVGYVAVGNDFDARLCGGADAQQRGA
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    orf34-1  LDVVLVEGDDFLYADGGADFLGNLRLLFFGGEDAHNVGYVAVGNDFDARLCGGADAQQRGA
      240     250     260     270     280     290

      310     320     330     340     350     360
    40  orf34a.pep  DFGCVPSVAGDVAGSARQGGDGNVXVHAFGGLFGTCNLTD ELFLAFGGDLSEQQVAVVA
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    orf34-1  DFGCVPSVAGDVAGSARQGGDGNIVVHAFGGLFGTCNLTD ELFFAFGGDLSEQQVAVVA
      300     310     320     330     340     350

      370     380     390     400     410     420
    45  orf34a.pep  DNGDLGRVXFGLVLAQIGAGGGFDTQRHYVVGXRAGGS AVDGGFRADRRRAADDCADAA
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    orf34-1  DDGDLGRVAFGLVLAQIGTGGGFDTQRHNVVGLRAGGS AVDGGFRADGGASDYCADAA
      360     370     380     390     400     410

      430     440     450     460
    50  orf34a.pep  AEGKAEDGGSQGADGVRFGFHRVLPFLGVSDGIALRHAVX
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
    orf34-1  AKGKAENGGNQADGVRFGFHRVLPFLGVSDGIALRHAVX
      420     430     440     450
  
```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF34 shows 77.6% identity over a 161aa overlap with a predicted ORF (ORF34.ng) from *N. gonorrhoeae*:

```

    orf34.pep  QKSLSRISLWGLGGVFFGVSGLVWFSLVGSXE-----CAC  35
  
```



-163-

```

|| |||||:|||||
orf34ng      MMMPFIMLPWIAGVPAVPGQKRLSRISLWGLAGVFFGVSGLVWFSLVGSFSLGVSLGCAC 60
5 orf34.pep  FSGVSFRGSGRGTfVGSTGVSLSVFSACVXGVVRLPVGLSCV-----GRLXXLTRFFLGA 90
orf34ng      FSGVSFRGSGWGAfVGSTGVSLSVFSACVP----VPVNESAAARAASEGR--GLTRFFLGA 114
orf34.pep    AGDVILLPLSSVPSGCGSDEAAWCSGWAASCPTTTFPGSQNSVSRGLSVCCGSAXRVLS 150
10 orf34ng    AGDGSPLPLSSVPSGCGSDEAAWCSGWAASCPTAFPGSQNSVSRGLSVCCGSVWRVLS 174
orf34.pep    S 175
orf34ng      PFGLNVLTMPTANAPMAVIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSD 234

```

15 The complete length ORF34ng nucleotide sequence <SEQ ID 213> is:

```

1 ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGCGGGTG TGCCTGCCGT
51 GCCGGGTCAA AAGAGGTTGT CGAGAATCTC TTTATGGGGT TTGGCCGGCG
101 TGTTTTTCGG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTTTT
151 TCTTTGGGTG TTTCTTTGGG CTGCGCCTGT TTTTCGGGTG TTTCTTTTCG
20 201 GGGTTCGGGA TGGGGGGCGT TTGTGGGCAG TACGGGGGTT TCTTTGAGTG
251 TGTTTTCAGC TTGTGTTCCG GTGCCGGTTA ACGAATCGGC TGCCCGGGCC
301 GCATCCGAAG GGCGCGGTTT gACCCGGTTT TTCTTGGGTG CGGCAGGGGA
351 CGGCAGTCCG CTGCCGCTTT CTCTGTGCC GTCCGGCTGT GCGGGTTCGG
401 ATGAGGCGGC GTGGTGGTGT TCGGGTTGGG CGGCATCTTG TCCGACGGCG
25 451 CCGTTTGGCA GCCAGAATTC GGTTTCGCGG GGGCTGTCGG TGTGTTGCGG
501 TTCGGTTTGG AGGGTTTTGT CGCCGTTTCG GTTGAATGTG CTGACGATGC
551 CTACTGCCAA TGCGCCGATG GCGGTGATAC AGATGAGCAA TACGGCGCGT
601 ATCAGGAGTT TGGGGGTCAG CCTGAAGGGT TTGTTCCGGT TTTTTCGCAT
651 TTTGATTGTG CTTTGGGGT GTCGGGCAAT GCCGTCTGAA GCGGTTTCAG
30 701 ACGGCATTGC CGAGTCAGCG TTGGACGTAG TTTTGGTAGA GGGTAATGAC
751 TTTTGTACG CCGAcggTGG TGCTGACTTT TTGGGTAATC TGCGCCTGTT
801 CTTCCGGGGT GAGGATGCCC ATAACGTAGG TTACATTGCC GTAGGTAATG
851 ATTTTGACGC GCGCCTGTGT AGCGGGGCTG ATGCCAGCA GcgtgGCGCG
901 GACTTTGGAC GTGTTCCAAG TGTCGCCGCG GATGTCGCCG GCAGTGC CGCG
35 951 GCAGGGAGGC GACGTAATG TAGTTGTATA CGCCTTCGGC GGCCTGTTTCG
1001 GAACGTGCAA TCTGACCGAC GAACTGTTTT TCGCCTTCGG TGGCGACTTG
1051 TCCGAGCAGC AGCAGGTGGC GGTGTAGACC GACGACGGAG ATTTGGGGCG
1101 TGTAGCCTTT GGTTTGGTTG TTTTGGCGCA GGTAGGAACG GCGGTGGTTG
1151 TCGATACGCA ACGCCATAAC GTtgtCATCG GTTtgcgcgc CGGTGGTTcg
40 1201 gCGGTCGATG ACGGATTTTG CGCCGACGGC GGCCCGCCG ACGACTGCGC
1251 TGAAGCAGCC GCCGAGGGCA AGGCTGAGGA CGGCGGCAAT CAGGGTGCGG
1301 ACGGTGTGTG GTTTGGGTTT CATCGGGGAC TTCCTTTCTT GGGCGTTTCA
1351 GACGGCATTG CTTTGCGCCA TGCCGTCTGA

```

This encodes a protein having amino acid sequence <SEQ ID 214>:

```

45 1 MMMPFIMLPW IAGVPAVPGQ KRLSRISLWG LAGVFFGVSG LVWFSLVGSF
51 SLGVSLGCAC FSGVSFRGSG WGAfVGSTGV SLSVFSACVP VPVNESAAARA
101 ASEGRGLTRF FLGAAGDGSP LPLSSVPSCG AGSDEAAWVC SGWAASCPTA
151 PFGSQNSVSR GLSVCCGSVW RVLSPFGLNV LTMPTANAPM AVIQMSNTAR
50 201 IRSLGVSLKG LFGFFAILIV LLGCRAMPSE GGSDGIAESA LDVVVLVEGND
251 FLYADGGADF LGNLRLFFGG EDAHNVGYIA VGNDFDARLC SGADAQQRGA
301 DFGRVPSVAG DVAR SARQGG DGNVVVYAFG GLFGTCNLTD ELFFAFGGDL
351 SEQQQVAVVA DDGDLGRVAF GLVVLAQVGT GGGFDTQRHN VVIGLRAGGS
401 AVDDGFCADG GPADDCAEEA AEGKAEDGGN QGADGVWFGF HRGLPFLGVS
451 DGIALRHAV*

```

55 ORF34ng and ORF34-1 show 90.0% identity in 459 aa overlap:

```

10 20 30 40 4 50
orf34-1.pep  MMMPFIMLPWIAGVPAVPGQNRLSRISLWGLGGVFFGVSGLVWFSLVGS-----LGAC
60 orf34ng    MMMPFIMLPWIAGVPAVPGQKRLSRISLWGLAGVFFGVSGLVWFSLVGSFSLGVSLGCAC
10 20 30 40 50 60
orf34-1.pep  FSGVSFRGSGRGTfVGSTGVSLSVFSACVPASSGCLSVXAVSAGCGLTRFFLGAAGDGSP
65 orf34ng    FSGVSFRGSGWGAfVGSTGVSLSVFSACVPVPVNESAAARAASEGRGLTRFFLGAAGDGSP

```

-164-

		70	80	90	100	110	120
		120	130	140	150	160	170
5	orf34-1.pep	LPLSSVPSGCAGSDEAAWCSGWAASCTTTPFGSQNSVSRGLSVCCGSAXRVLSPFGLNV					
	orf34ng	LPLSSVPSGCAGSDEAAWCSGWAASCTTAPFGSQNSVSRGLSVCCGSVWRVLSPFGLNV					
		130	140	150	160	170	180
		180	190	200	210	220	230
10	orf34-1.pep	LTMPIANAPMAAIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA					
	orf34ng	LTMPTANAPMAVIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA					
		190	200	210	220	230	240
		240	250	260	270	280	290
15	orf34-1.pep	LDVVLVEGDDFLYADGGADFLGNLRLEFFGGEDAHNVGYVAVGNDFDARLCGGADAQQRGA					
	orf34ng	LDVVLVEGNDFLYADGGADFLGNLRLEFFGGEDAHNVGYIAVGNDFDARLCGSGADAQQRGA					
		250	260	270	280	290	300
		300	310	320	330	340	350
20	orf34-1.pep	DFGCVPSVAGDVAGSARQGGDGNIVVHAFGGLFGTCNLTDLFFAFGGDLSEQQQVAVVA					
	orf34ng	DFGRVPSVAGDVARSARQGGDGNVVVYAFGGGLFGTCNLTDLFFAFGGDLSEQQQVAVVA					
		310	320	330	340	350	360
		360	370	380	390	400	410
30	orf34-1.pep	DDGDLGRVAFGLVLVLAQIGTGGGFDTORHNVVVGRLRAGGSAVDGGFRADGGASDYCADAA					
	orf34ng	DDGDLGRVAFGLVLVLAQVGTGGGFDTORHNVVIGLRAGGSAVDDGFCADGGPADDCAEAA					
		370	380	390	400	410	420
		420	430	440	450		
35	orf34-1.pep	AKGKAENGGNQGADGVRFGRVLPFLGVSDGIALRHAVX					
	orf34ng	AEGKAEDGGNQGADGVWFGFRGLPFLGVSDGIALRHAVX					
		430	440	450	460		

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 26

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 215>:

```

45      1  ATGAAACCT  TCTTCAAAC  CCTTTCCGCC  GCCGCACTCG  CGCTCATCCT
      51  CGCCGCCTGC  GGATT.CAAA  AAGACAGCGC  GCCCGCCGCA  TCCGCTTCTG
     101  CCGCCGCCGA  CAACGGCGCG  GCGTAAAAAA  GAAATCGTCT  TCGGCACGAC
     151  CGTCGGCGAC  TTCGGCGATA  TGGTCAAAGA  ACAATCCAA  GCCGAGCTGG
     201  AGAAAAAGG  CTACACCGTC  AACTGGTTCG  AGTTTACCGA  CTATGTACGC
     251  CCGAATCTGG  CATGGCTGA  GGGCGAGTTG

```

50 This corresponds to the amino acid sequence <SEQ ID 216; ORF4>:

```

      1  MKTFFKTLA  AALALILAAC  G.QKDSAPAA  SASAADNGA  AKKEIVFGTT
     51  VGDFGDMVKE  QIQAELEKKG  YTVKLVEFTD  YVRPNLALAE  GEL

```

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 217>:

```

55      1  ATGAAACCT  TCTTCAAAC  CCTTTCCGCC  GCCGCACTCG  CGCTCATCCT
      51  CGCCGCCTGC  GCGGTCAAA  AAGACAGCGC  GCCCGCCGCA  TCCGCTTCTG
     101  CCGCCGCCGA  CAACGGCGCG  GCGAAAAAAG  AAATCGTCTT  CGGCACGACC
     151  GTCGCGGACT  TCGGCGATAT  GGTCAAAGAA  CAAATCCAA  CCGAGCTGGA
     201  GAAAAAAGGC  TACACCGTCA  AACTGGTTCG  GTTTACCGAC  TATGTACGCC

```

-165-

251 CGAATCTGGC ATTGGCTGAG GCGGAGTTGG ACATCAACGT CTTCCAACAC  
 301 AAACCTATC TTGACGACTT CAAAAAGAA CACAATCTGG ACATCACCAG  
 351 AGTCTTCCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA  
 401 AATCGCTGGA AGAAGTCAAA GACGGCAGCA CCGTATCCGC GCCCAACGAC  
 451 CCGTCCAACT TCGCCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTGGAT  
 501 CAAACTCAAA GACGGCATCA ATCCGTTGAC CGCATCCAAA GCGGACATCG  
 551 CCGAGAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CGCGCAACTG  
 601 CCGCGTAGCC GCGCCGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC  
 651 CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT  
 701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA  
 751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA  
 801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGAATGAAG  
 851 GCGCAGCCAA ATAA

This corresponds to the amino acid sequence <SEQ ID 218; ORF4-1>:

15 1 MKTFFKTLA AALALILAAC GGQKDSAPAA SASAAADNGA AKKEIVFGTT  
 51 VGDFGDMVKE QIQAELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFQH  
 101 KPYLDDEFKE HNLDTVEVFQ VPTAPLGLYP GKLSLEEVEK DGSTVSAPND  
 151 PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLNKI KIVELEAAQL  
 201 PRSRADVDEFA VVNGNYAISS GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ  
 251 WLKDVTEAYN SDAFKAYAHK RFEQYKSPAA WNEGAAK\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF4 shows 93.5% identity over a 93aa overlap with an ORF (ORF4a) from strain A of *N. meningitidis*:

25	orf4.pep	10	20	30	40	50	59
	orf4a	10	20	30	40	50	60
30	orf4.pep	60	70	80	90		
35	orf4a	70	80	90	100	110	120
	orf4a	130	140	150	160	170	180

The complete length ORF4a nucleotide sequence <SEQ ID 219> is:

40 1 ATGAAAACCT TCTTCAAAAC CCTTCCGCC GCCGCACTCG CGCTCATCCT  
 51 CGCCGCGTGC GGCGGTCAAA AAGATAGCGC GCCCGCCGCA TCCGCTTCTG  
 101 CCGCCGCGCA CAACGGCGCG CGGAANAAAG AAATCGTCTT CGGCACGACC  
 151 GTCGGCGACT TCGGCGATAT GGTCAAAGAA CANATCCAAC CCGAGCTGGA  
 201 GAAAAAAGGC TACACCGTCA AACTGGTCGA GTNTACCGAC TATGTGCGCN  
 45 251 CGAATCTGGC ATTGGCTGAG GCGGAGTTGG ACATCAACGT CTTNCAACAC  
 301 ANACNCTATC TTGACGACTN CAAAAANAA CACAATCTGG ACATCACCNN  
 351 AGTCTTNCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA  
 401 AATCGCTGGA NNAAGTCAAA GANGGCAGCA CCGTATCCGC GCCCAACGAC  
 451 CCGTNNNACT TCGNCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTNGAT  
 50 501 CAAACTCAAA GACNGCATCA NNNNGNNGNN NNNANCNANA NNNGANANN  
 551 NNNNANNNT NNNNNNNNN NNNNNCNCN NNNNNNNNN NNNNNNNNN  
 601 NCGNNTNNNN NNGCNNNNNT NNANNNTNN NNCNNCNNNN NNNNTNNNN  
 651 NANNANNAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT  
 701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA  
 55 751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA  
 801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGAATGAAG  
 851 GCGCAGCCAA ATAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 220>:

1 MKTFFKTLA AALALILAAC GGQKDSAPAA SASAAADNGA AXKEIVFGTT

-166-

51 VGDFGDMVKE XIQPELEKKG YTVKLVEFTD YVRXNLALAE GELDINXQH  
 101 XXYLDDXKKX HNLDITXVXQ VPTAPLGLYP GKLSLXXVK XGSTVSAPND  
 151 PXXFXRVLM LDELGXIKLK DXIXXXXXXX XXXXXXXXXX XXXXXXXXXX  
 201 XXXXAXXXXX XXXXXXXXXX GMKLTEALFQ EPSFAYVNS AVKTADKDSQ  
 251 WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK\*

A leader peptide is underlined.

Further analysis of these strain A sequences revealed the complete DNA sequence <SEQ ID 221>:

1 ATGAAAACCT TCTTCAAAAC CCTTTCCGCC GCCGCACTCG CGCTCATCCT  
 51 CGCCGCTGC GCGGCTCAA AAGATAGCGC GCCGCCGCA TCCGCTTCTG  
 101 CCGCGCCGA CAACGGCGCG GCGAAAAAG AAATCGTCTT CGGCACGACC  
 151 GTCGCGACT TCGCGATAT GGTCAAAGAA CAAATCCAAC CCGAGCTGGA  
 201 GAAAAAGGC TACACCGTCA AACTGGTCGA GTTTACCGAC TATGTGCGCC  
 251 CGAATCTGGC ATGGCTGAG GCGAGTTGG ACATCAACGT CTTCCAACAC  
 301 AAACCTATC TTGACGACTT CAAAAAGAA CACAATCTGG ACATCACCAG  
 351 AGTCTTCAA GTCCGACCG CGCCTTTGGG ACTGTACCGG GGCAAGCTGA  
 401 AATCGCTGGA AGAAGTCAA GACGGCAGCA CCGTATCCGC GCCCAACGAC  
 451 CCGTCCAAT TCGCCGCGT CTGGTGATG CTCGACGAAC TGGGTGGAT  
 501 CAAACTCAA GACGGCATCA ATCCGCTGAC CGCATCCAA GCGGACATTG  
 551 CCGAAAACCT GAAAAATC AAAATCGTCG AGCTTGAAGC CGCGCAACTG  
 601 CCGCGTAGCC GCGCGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC  
 651 CATAGCAGC GGATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT  
 701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA  
 751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA  
 801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGAATGAAG  
 851 GCGCAGCCAA ATAA

This encodes a protein having amino acid sequence <SEQ ID 222; ORF4a-1>:

1 MKTFFKTLA AALALILAAC GGQKDSAPAA SASAADNGA AKKEIVFGTT  
 51 VGDFGDMVKE XIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINXQH  
 101 KPYLDDFKKE HNLDITEVFO VPTAPLGLYP GKLSLEEVK DGSTVSAPND  
 151 PSNFARVLM LDELGWIKLK DGINPLTASK ADIAENLKNI KIVELEAAQL  
 201 PRSRADVDF VVNGNYAISS GMKLTEALFQ EPSFAYVNS AVKTADKDSQ  
 251 WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK\*

ORF4a-1 and ORF4-1 show 99.7% identity in 287 aa overlap:

35	orf4a-1	10	20	30	40	50	60
		MKTFFKTLA	AALALILAAC	GGQKDSAPAA	SASAADNGA	AKKEIVFGTT	VGDFGDMVKE
	orf4-1	MKTFFKTLA	AALALILAAC	GGQKDSAPAA	SASAADNGA	AKKEIVFGTT	VGDFGDMVKE
40	orf4a-1	70	80	90	100	110	120
		QIQPELEKKG	YTVKLVEFTD	YVRPNLALAE	GELDINXQH	KPYLDDFKKE	HNLDITEVFO
	orf4-1	QIQAELEKKG	YTVKLVEFTD	YVRPNLALAE	GELDINXQH	KPYLDDFKKE	HNLDITEVFO
45	orf4a-1	130	140	150	160	170	180
		VPTAPLGLYP	GKLSLEEVK	DGSTVSAPND	PSNFARVLM	LDELGWIKLK	DGINPLTASK
	orf4-1	VPTAPLGLYP	GKLSLEEVK	DGSTVSAPND	PSNFARVLM	LDELGWIKLK	DGINPLTASK
50	orf4a-1	190	200	210	220	230	240
		ADIAENLKN	IKIVELEAAQL	PRSRADVDF	AVVNGNYAIS	SGMKLTEALF	QEPSFAYVNS
	orf4-1	ADIAENLKN	IKIVELEAAQL	PRSRADVDF	AVVNGNYAIS	SGMKLTEALF	QEPSFAYVNS
55	orf4a-1	250	260	270	280		
		AVKTADKDSQ	WLKDVTEAYN	SDAFKAYAHK	RFEGYKSPAA	WNEGAAKX	
	orf4-1	AVKTADKDSQ	WLKDVTEAYN	SDAFKAYAHK	RFEGYKSPAA	WNEGAAKX	
60		250	260	270	280		

Homology with an outer membrane protein of *Pasteurella haemolítica* (accession q08869).

ORF4 and this outer membrane protein show 33% aa identity in 91aa overlap:

5	lip2.pasha																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
---	------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

## 20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF4 shows 93.6% identity over a 94aa overlap with a predicted ORF (ORF4.ng) from *N. gonorrhoeae*:

						10	20	30			
25	orf4nm.pep					MKTFFK	TL	SAAALALILAACGXQKDSAPAA			
	orf4ng	RANAVX	TNP	DGR	TPCL	SFL	FETATTSGEN	MKTFFK	TL	STAS	LALILAACGGQKDSAPAA
		200	210	220	230	240	250				
30	orf4nm.pep		40	50	60	70	80	89			
	orf4ng	SASA-	AADNGAAKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVEFTDYVRPNLALA								
		260	270	280	290	300	310				
35	orf4nm.pep	90	EGEL								
	orf4ng	EGEL	DIN	VQHKPYLDDFKKEHNLDITEAFQVPTAPLGLYPGKLKSLEEVKDGSTVSAPN							
		320	330	340	350	360	370				

40 The complete length ORF4ng nucleotide sequence <SEQ ID 223> was predicted to encode a protein having amino acid sequence <SEQ ID 224>:

45	1	MKTFFKTLST	ASLALILAAC	GGQKDSAPAA	SAAAPSAADNG	AAKKEIVFGT
	51	TVGDFGDMVK	EQIQAELEKK	GYTVKLVEFT	DYVRPNLALA	EGELDINVQ
	101	HKPYLDDFKK	EHNL DITEAF	QVPTAPLGLY	PGKLKSLEEV	KDGSTVSAPN
	151	DPSNFARALV	MLNELGWIKL	KDGINPLTAS	KADIAENLKN	IKIVELEAAQ
	201	LPRSRADVDF	AVVNGNYAIS	SGMKLTEALF	QEPSFAYVNW	SAVKTADKDS
	251	QWLKDVTEAY	NSDAFKAYAH	KRFEGYKYP	AWNEGAAK*	

Further analysis revealed the complete length ORF4ng DNA sequence <SEQ ID 225> to be:

50	1	atgAAAACCT	TCTTCAAAAC	cctttccgcc	gccgcaCTCG	CGCTCATCCT
	51	CGCAGCCTGc	ggCggtcaAA	AAGACAGCGC	GCCCgcagcc	tctgcCGCCG
	101	CCCCTTCTGC	CGATAACGgc	gCgGCGAAAA	AAGAAAtcgt	ctTCGGCAGC
	151	Accgtggcg	acttcggcgA	TAtggTCAAA	GAACAAATCC	AagcCGAgct
55	201	gGAGAAAAAA	GgctACACcg	tcAAattggt	cgaatttacc	gactatgtGC
	251	gCCCGAATCT	GGCATTGGCG	GAGGGCGAGT	TGGACATCAA	CGTCTTCCAA
	301	CACAAACCCT	ATCTTGACGA	TTTCAAAAAA	GAACACAACC	TGGACATCAC
	351	CGAAGCCTTC	CAAGTGCCGA	CCGCGCCTTT	GGGACTGTAT	CCGGGCAAAC
60	401	TGAAATCGCT	GGAAGAAGTC	AAAGACGGCA	GCACCGTATC	CGCGCCCAac
	451	gACccgTCCA	ACTTCGCACG	CGCCTTGGTG	ATGCTGAACG	AACTGGGTTG
	501	GATCAAACTC	AAAGACGGCA	TCAATCCGCT	GACCGCATCC	AAAGCCGACA
	551	TCGCGGAAAA	CCTGAAAAAC	ATCAAAATCG	TCGAGCTTGA	AGCCGCACAA



	lip2_pasha	TEVAVKIAKEKYGLDVELVQFTEYTPQNAALHSKDLDNAFQTVTPYLEQEVKDRGYKLAI	60	70	80	90	100	110
5	orf4ng-1.pep	120 130 140 150 160 170 AFQVPTAPLGLYPGKLKLSLEEVKDGSTVSAFNDPSNFARALVMLNELGWIKLKDGINPLT :::  ::  :::  :::  ::  ::  :::  ::  :::  ::  ::						
	lip2_pasha	IGNTLVWPPIAAYSKKIKNISELKDGATVAIPNNASNTARALLLQAHGLLKLKDKPN-VF	120	130	140	150	160	170
10	orf4ng-1.pep	180 190 200 210 220 230 ASKADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTE--ALFQEPSFA  ::            :: ::       ::   :::   ::: : : : :						
	lip2_pasha	ATENDIENPNKNIKIVQADTSLLRMLDDVELAVINNTYAGQAGLSPDKDGIIVESKDSP	180	190	200	210	220	230
15	orf4ng-1.pep	240 250 260 270 280 289 YVNWSAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKYPAAWNEGAAX     : : :  :  : ::::      ::						
	lip2_pasha	YVNLVVSREDNKDDPRLQTFVKSFQTEEVFQEALKLFNGGVVKGW	240	250	260	270		
20			240	250	260	270		

Based on this analysis, including the homology with the outer membrane protein of *Pasteurella haemolitica*, and on the presence of a putative prokaryotic membrane lipoprotein lipid attachment site in the gonococcal protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF4-1 (30kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 8A and 8B show, respectively, the results of affinity purification of the His-fusion and GST-fusion proteins. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), Western blot (Figure 8C), FACS analysis (Figure 8D), and a bactericidal assay (Figure 8E). These experiments confirm that ORF4-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 8F shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF4-1.

### Example 27

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 227>:

	1	CCTCGTCGTC	CTCGGCATGC	TCCAGTTTCA	AGGGGCGATT	TACTCCAAGG
	51	CGGTGGAACG	TATGCTCGGC	ACGGTCATCG	GGCTGGGCGC	GGGTTTGGGC
40	101	GTTTTATGGC	TGAACCAGCA	TTATTTCCAC	GGCAACCTCG	TCTTCTACCT
	151	CACCGTCGGC	ACGGCAAGCG	CAC TGCGCGG	CTGGGCGCGC	GTCGGCAAAA
	201	ACGGCTACGT	CCCTmTGCTG	GCAGGGCTGA	CGATGTGTAT	GCTCATCGGC
	251	GACAACGGCA	GCGAATGGCT	CGACAGCGGA	CTCATGCGCG	CCATGAACGT
	301	CCTCATCGGC	GyGGCCATCG	CCATCGCCCG	CGCCAAACTG	CTGCCGCTGC
	351	AATCCACACT	GATGTGGCGT	TTCATGCTTG	CCGACAACCT	GGCCGACTGC
45	401	AGCAAAATGA	TTGCCGAAAT	CAGCAACGGC	AGGCGCATGA	CCCGCGAACG
	451	CCTCGAGGAG	AACATGGCGA	AAATGCGCCA	AATCAACGCA	CGCATGGTCA
	501	AAAGCCGCAG	CCATCTCGCC	GCCACATCGG	GCGAAAGCTG	CATCAGCCCC
	551	GCCATGATGG	AAGCCATGCA	GCACGCCAC	CGTAAATCG	TCAACACCAC
	601	CGAGCTGCTC	CTGACCACCG	CCGCCAAGCT	GCAATCTCCC	AAACTCAACG

5  
 651 GCAGCGAAAT CCGGCTGCTT GACCGCCACT TCACACTGCT CCAAAC....  
 701 ..... GC AGACACGCCC GCCGCATCCG  
 751 CATCGACACC GCCATCAACC CCGAACTGGA AGCCCTCGCC GAACACCTCC  
 801 ACTACCAATG GCAGGGCTTC CTCTGGCTCA GCACCGATAT GCGTCAGGAA  
 851 ATTTCCGCCC TCGTCATCCT GCTGCAACGC ACCCGCCGCA AATGGCTGGA  
 901 TGCCACGAA CGCCAACACC TGCGCCAAAG CCTGCTTGA

This corresponds to the amino acid sequence <SEQ ID 228; ORF8>:

10  
 1 .....PRRP RHAPVSRGDL LQGGGYARH GHRAGRGFGR FMAEPALFPR  
 51 QPPLLPHRRH GKRTGRLGGG RQKRLRPXAG RADDVYAHRR QRQRMARQRT  
 101 HARHERPHRR GHRHRRRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ  
 151 AHDPRTPRGE HGENAPNORT HGQKPQPSRR HIGRKLHQPR HDGSHAARPP  
 201 XNRQHHRAAP DHRRQAASQ TQRQNPAAQ PPLHTAPN.. .....Q  
 251 TRPPHPHRRH HQPRTGSPRR TPPLPMAGLP LAQHRYASGN FRPRHPAATH  
 301 PPQMAGCPRT PTPAPKPA\*

15 Computer analysis of this amino acid sequence gave the following results:

#### Sequence motifs

ORF8 is proline-rich and has a distribution of proline residues consistent with a surface localization. Furthermore the presence of an RGD motif may indicate a possible role in bacterial adhesion events.

#### 20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF8 shows 86.5% identity over a 312aa overlap with a predicted ORF (ORF8.ng) from *N. gonorrhoeae*:

25  
 orf8ng 1 MDRDDLRRPRHAPVPRRDLQGGGYARYGHRAGRGFGRFMAEPALFPR 50  
 orf8.pep 1 .....PRRPRHAPVSRGDLQGGGYARHGHRAGRGFGRFMAEPALFPR 44  
 30  
 orf8ng 51 QPILLEDHRHGKRTGRLGGGRQKRLRPYVGGADDVHAHRRQRQRMARQRP 100  
 orf8.pep 45 QPPLLPHRRHGKRTGRLGGGRQKRLRPXAGRADDVYAHRRQRQRMARQRT 94  
 35  
 orf8ng 101 DARDERPHRRRHRCRRQTAAAEIHTDVAFHACRQPGRLQNDNCRNQQRQ 150  
 orf8.pep 95 HARHERPHRRGHRHRRRQTA AEIHTDVAFHACRQPGRMQNDNCRNQQRQ 144  
 40  
 orf8ng 151 AYDARTFGAEYQONAPNORTHGQKPQPPRRHIGRKPHQPLHDGSHAARPP 200  
 orf8.pep 145 AHDPRTPRGEHGENAPNORTHGQKPQPSRRHIGRKLHQPRHDGSHAARPP 194  
 45  
 orf8ng 201 QNRQHHRAAPDHRRQAASQTQRQNPAAQ PPLHTAPNRPATNRRPHQRQ 250  
 orf8.pep 195 XNRQHHRAAPDHRRQAASQTQRQNPAAQ PPLHTAPN.....Q 244  
 orf8ng 251 TRPPHPHRRHQPRTGSPRRTPPLPMAGFPLAQHQYASGNFRPRHPPATH 300  
 orf8.pep 245 TRPPHPHRRHQPRTGSPRRTPPLPMAGLPLAQHRYASGNFRPRHPAATH 294  
 orf8ng 301 PPQMAGCPRTPTPAPKPA\* 319  
 orf8.pep 295 PPQMAGCPRTPTPAPKPA\* 313

50 The complete length ORF8ng nucleotide sequence <SEQ ID 229> is predicted to encode a protein having amino acid sequence <SEQ ID 230>:

55  
 1 MDRDDLRRP RHAPVPRDL LQGGGYARY GHRAGRGFGR FMAEPALFPR  
 51 QPILLPDHRR GKRTGRLGGG RQKRLRPYVG GADDVHAHRR QRQRMARQRP  
 101 DARDERPHRR RHRHRCRRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ  
 151 AYDARTFGAE YQONAPNORT HGQKPQPPRR HIGRKPHQPL HDGSHAARPP



201 QNRQHHRAAP DHRROAAISQ TORQRNPAAR PPLHTAPNRP ATNRRPHORO  
 251 TRPPHPHRHR HQPRTGSPRR TPPLPMAGFP LAQHQYASGN FRPRHPPATH  
 301 PPQMAGCPRT PTPAPKPA\*

Based on the sequence motifs in these proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 28

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 231>:

```

1  ..GAAATCAGCC TGCGGTCCGA CNACAGGCCG GTTTCCTGN CGAAGCGGCG
10  51  GGATTCGGAA CGTTTTCTGC TGTTGGACGG CGGCAACAGC CGGCTCAAGT
    101  GGGCGTGGGT GGA AACGCGC ACCTTCGCAA CCGTCGGTAG CGCGCCGTAC
    151  CGCGATTTGT CGCCTTTGGG CGCGGAGTGG GCGGAAAAGG CGGATGGAAA
    201  TGTCCGCATC GTCGGTTGCG CTGTGTGCGG AGAATTCAAA AAGGCACAAG
    251  TGCAGGAACA GCTCGCCCGA AAAATCGAGT GGCTGCCGTC TTCCGCACAG
15  301  GCTTT.GGCA TACGCAACCA CTACCGCCAC CCCGAAGAAC ACGGTTCCGA
    351  CCGCTGGTTC AACGCCTTGG GCAGCCGCCG CTTACGCCGC AACGCCTGCG
    401  TCGTCGTCAG TTGCGGCACG GCGGTAACGG TTGACGCGCT CACCGATGAC
    451  GGACATTATC TCGGAGA.GG AACCATCATG CCCGTTTCC ACCTGATGAA
    501  AGAATCGCTC GCCGTCCGAA CCGCCAACT CAACCGGCAC GCCGGTAAGC
20  551  GTTATCCTTT CCCGACCGG..

```

This corresponds to the amino acid sequence <SEQ ID 232; ORF61>:

```

1  ..EISLRSDXRP VSVXKRRDSE RFLLLDGGNS RLKWAVVENG TFATVGSAPY
25  51  RDLSPGLAEW AEKADGNVRI VCAVCGEFK KAQVQEQLAR KIEWLPSSAQ
    101  AXGIRNHYRH PEEHGSDFWF NALGSRFRFSR NACVVVSCGT AVTVDALTD
    151  GHYLGXGTIM PGFHLMKESL AVRTANLNHRH AGKRYPFPT..

```

Further work revealed the complete nucleotide sequence <SEQ ID 233>:

```

1  ATGACGGTTT TGAAGCTTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA
30  51  CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
    101  CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CATACGCGGG
    151  CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTGGCGGT
    201  TTTGATGCC GAAGTTTTGC GCGAGCTGGG GGAAAGGTCT GGTTTTCAGA
    251  CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
    301  CGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGCG TGACCCACT
    351  GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
15  401  GCGAGTGTCT GATGTTCACT TTTGGCTGGG TGTTTGACCG GCCGCGATAT
    451  GAGTTGGGTT CGCTGTCGCC GTTGCGGCA GTGGCGTCTC GGCGCGCCTT
    501  TCGCGTTTA GGTTTGGATG TGCAGATTAA GTGGCCCAAT GATTGTGTTG
    551  TCGGACGCGA CAAATTGGGC GCATTCTGA TTGAAACGGT CAGGACGGGC
    601  GGCAAAACGG TTGCCGTGGT CCGTATCGGC ATCAATTTTG TCCTGCCCAA
40  651  GGAAGTAGAA AATGCCGCTT CCGTGCAATC GCTGTTTCAG ACGGCATCGC
    701  GGCGGGGCAA TGCCGATGCC GCCGTGCTGC TGGAAACGCT GTTGGTGGAA
    751  CTGGACGCGG TGTGTTGCA ATATGCGCGG GACGGATTG CGCCTTTTGT
    801  GCGGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
    851  TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTTAAAGG CGTGGACGGA
15  901  CAAGGCGTTT TGCATTGGA AACGGCAGAG GGCAACAGA CGGTCGTCAG
    951  CGGCGAAATC AGCCTGCGGT CCGACGACAG GCCGGTTTCC GTGCCGAAGC
20  1001  GGCGGGATTG GGAACGTTTT CTGCTGTTGG ACGGCGGCAA CAGCCGGCTC
    1051  AAGTGGCGT GGGTGAAAAA CGGCACGTTT GCAACCGTCG GTAGCGCGCC
    1101  GTACCGGAT TTGTGCGCTT TGGGCGCGGA GTGGGCGGAA AAGGCGGATG
50  1151  GAAATGTCCG CATCGTCGGT TGCCTGTGT GCGGAGAATT CAAAAAGGCA
    1201  CAAGTCGAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
    1251  ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT
    1301  CCGACCGCTG GTTCAACGCC TTGGGACGCC GCGGCTTCAG CCGCAACGCC
    1351  TCGCTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
55  1401  TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTTC CACCTGATGA
    1451  AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACGGGCA CGCCGGTAAG
    1501  CGTTATCCTT TCCCGACCAC AACGGGCAAT GCCGTGCGCA GCGGCATGAT
    1551  GGATGCGGTT TGCGGCTCGG TTATGATGAT GCACGGGCGT TTGAAAGAAA
    1601  AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG

```

1651 GCAAAAGTTG CCGAAGCCCT GCCGCCTGCA TTTTGGCGG AAAATACCGT  
1701 GCGCGTGGCG GACAACCTCG TCATTTACGG GTTGTGTAAC ATGATTGCCG  
1751 CCGAAGGCAG GGAATATGAA CATATTTAA

This corresponds to the amino acid sequence <SEQ ID 234; ORF61-1>:

5	1	MTVLKLSHWR	VLAELADGLP	QHSVQLARMA	DMKPOQLNGF	WQOMPAHIRG
	51	LLRQHDGYWR	LVRPLAVFDA	EGLRELGERS	GFQTALKHEC	ASSNDEILEL
	101	ARIAPDKAHK	TICVTHLQSK	GRGRQGRKWS	HLRGECLMFS	FGWVFRDPQY
	151	ELGSLSPVAA	VACRRALSRL	GLDVQIKWPN	DLVVGRDKLG	GILIETVRTG
	201	GKTVAVVGIG	INFVLPKEVE	NAASVQSLFQ	TASRRGNADA	AVLLETLLVE
10	251	LDAVLLQYAR	DGFAPFVAEY	QAANRDHGKA	VLLLRDGETV	FEGTVKGVDG
	301	QGV LHLETAE	GKQTVVSGEI	SLRSSDRPVS	VPKRRDSERF	LLLDGGNSRL
	351	KWAWVENGTF	ATVGSAPYRD	LSPLGAEWAE	KADGNVRIVG	CAVCGEFKKA
	401	QVQEQLARKI	EWLPSSAQAL	GIRNHYRHPE	EHGSDRWENA	LGSRRFSRNA
	451	CVVVSCGTAV	TVDALTDGDH	YLGGTIMPGF	HLMKESLAVR	TANLNRHAGK
15	501	RYPFPTTTGN	AVASGMDAV	CGSVMMMHGR	LKEKTGAGKP	VDVITGGGA
	551	AKVAEALPPA	FLAENTVRVA	DNLVIYGLLN	MIAAEGREYE	HI*

Figure 9 shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF61-1. Further computer analysis of this amino acid sequence gave the following results:

Homology with the baf protein of *B. pertussis* (accession number U12020).

20 ORF61 and baf protein show 33% aa identity in 166aa overlap:

	orf61	23	LLLDGGNSRLKWAWVE-NGTFATVGSAPYR----DLSPLGAWEAEKADGNVRIVGCAVCG	77
			+L+D GNSRLK W + + A AP DL LG A R +G V G	
	baf	3	ILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLDALGRWLATLPRRPQRALGVNVAG	62
25	orf61	78	EFKKAQVQEQLAR---KIEWLPSSAQAXGIRNHYRHPEEHGSDRW---FNALGSRRFSRN	131
			+ + L I WL + A G+RN YR+P++ G+DRW L +	
	baf	63	LARGEAIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGADRWACMVGVLARQPSVHP	122
30	orf61	132	ACVVVSCGTAVTVDALTDGHHYLGXGTIMPGFHLMKESLAVRTANL	177
			+V S GTA T+D + D + G G I+PG +M+ +LA TA+L	
	baf	123	PLLVASFGTATTLDITGPDNVFPG-GLILPGPAMMRGALAYGTAHL	167

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF61 shows 97.4% identity over a 189aa overlap with an ORF (ORF61a) from strain A of *N.*

35 *meningitidis:*

	orf61.pep	EISLRSDXRPVSVXKKRDSERFLLLDGGNS	10	20	30			
40	orf61a	TVFEGTVKGVDDGQVLHLETAEGKQT VVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNS	290	300	310	320	330	340
	orf61.pep	RLKWAWVENGT FATVGSAPYRDLSPLGAEWA EKADGNVRIVGC AVCGEFKKAQVQEQLAR	40	50	60	70	80	90
		:						
45	orf61a	RLKWAWVENGT FATVGSAPYRDLSPLGAEWA EKVDGNVRIVGC AVCGEFKKAQVQEQLAR	350	360	370	380	390	400
	orf61.pep	KIEWLPSSAQAXGIRNHYRHPEEHGSDRWFNALGSRFR SRNACVVVSCGTAVTVDALTD	100	110	120	130	140	150
50	orf61a	KIEWLPSSA QALGIRNHYRHPEEHGSDRWFNALGSRFR SRNACVVVSCGTAVTVDALTD	410	420	430	440	450	460
	orf61.pep	GHYLGXGTIMPGFHLMKESLAVRTANLN RHAGKRYFPFT	160	170	180	189		
	orf61a	GHYLG-GTIPMGFHLMKESLAVRTANLN RHAGKRYFPFTTTGNAVASGMMDAVCGSVM	470	480	490	500	510	520
60	orf61a	HGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHG LNLIAAEGG						

530                      540                      550                      560                      570                      580

The complete length ORF61a nucleotide sequence <SEQ ID 235> is:

```

1  ATGACGGTTT TGAAGCCTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA
5  51  CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
   101  CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CATACGCGGG
   151  CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTGGCGGT
   201  TTTCGATGCC GAAGGTTTGC GCGAGCTGGG GGAAAGGTCG GGTTTTCAGA
   251  CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
  10  301  GCGCGGATTG CGCCGGACAA GCGGCACAAA ACCATATGTG TGACCCACCT
   351  GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
   401  GCGAGTGTCT GATGTTTCAGT TTTGGCTGGG TGTTTGACCG GCCGCAGTAT
   451  GAGTTGGGTT CGCTGTCGCC TGTTGCGGCA GTGGCGTGCC GCGCGCCTT
   501  TCGCGGTTTG GGTTTGAAAA CGCAAATCAA GTGGCCAAAC GATTTGCTCG
  15  551  TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACGGT CAGGACGGGC
   601  GGCAAACCGG TTGCCGTGGT CCGTATCGGC ATCAATTTTCG TGCTGCCCAA
   651  GGAAGTGAA AACGCCGCTT CCGTGCAATC GCTGTTTCAG ACGGCATCGC
   701  GGCGGGGAAA TGCCGATGCC GCCGTGTTGC TGGAAACGCT GTTGGCGGAA
   751  CTTGATGCGG TGTGTTTGCA ATATGCGCGG GACGGATTTCG CGCCTTTTGT
  20  801  GCGCGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
   851  TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTTAAAGG CGTGGACGGA
   901  CAAGGCGTTC TGCATTGGA AACGGCAGAG GGCAAACAGA CGGTCGTCAG
   951  CGGCGAAATC AGCCTGCGGT CCGACGACAG GCCGGTTTCC GTGCCGAAGC
  1001  GCGCGGATTC GGAACGTTTT CTGCTGTTGG ACGGCGGCAA CAGCCGGCTC
  1051  AAGTGGCGGT GGGTGAAAAA CGGCACGTTT GCAACCGTCG GTAGCGCGCC
  25  1101  GTACCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGGTGGATG
   1151  GAAATGTCCG CATCGTCGGT TGCGCCGTGT GCGGAGAATT CAAAAAGGCA
   1201  CAAGTGACAG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
   1251  ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT
   1301  CCGACCGCTG GTTCAACGCC TTGGGACGCC GCCGCTTCAG CCGCAACGCC
  30  1351  TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
   1401  TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTTC CACCTGATGA
   1451  AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGGCA CGCCGGTAAG
   1501  CGTTATCCTT TCCCGACCAC AACGGGCAAT GCCGTCGCCA GCGGCATGAT
   1551  GGATGCGGTT TGCGGTCGGT TTATGATGAT GCACGGGCGT TTGAAAGAAA
  35  1601  AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG
   1651  GCAAAAGTTG CCGAAGCCCT GCCGCTGCA TTTTGGCGG AAAATACCGT
   1701  GCGCGTGGCG GACAACCTCG TCATTACCGG GCTGCTGAAC CTGATTGCCG
   1751  CCGAAGGCGG GGAATCGGAA CATACTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 236>:

```

40  1  MTVLKPSHWR VLAELADGLP QHVSQARMAD MKPQQLNGF WQOMPAHIRG
   51  LLRQHDGYWR LVRPLAVFDA EGLRELTERS GFQTLKHEC ASSNDEILEL
   101  ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWVFDPRQY
   151  ELGSLSPVAA VACRRALSRL GLKTQIKWPN DLVVGRDKLG GILIETVRTG
  45  201  GKTAVAVVIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLAE
   251  LDAVLLQYAR DGFAPFVAEY QAANRDHGKA VLLLRDGETV FEGTVKGVDG
   301  QGVHLLETAE GKQTVVSGEI SLRSDDRPVS VPKRRDSERF LLLDGGNSRL
   351  KWAVVENGTG ATVGSAFYRD LSPLGAWEAE KVDGNVRIVG CAVCGEFKKA
   401  QVQEQALARKI EWLPSSAQAL GIRNHYRHPE EHGSDRWFNA LGSRRFSRNA
   451  CVVVSCGTAV TVDALTDGHH YLGGTIMPFG HLMKESLAVR TANLNRHAGK
  50  501  RYPFPTTGN AVASGMMDAV CGSVMMHGR LKEKTGAGKP VDVIIITGGGA
   551  AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HT*

```

ORF61a and ORF61-1 show 98.5% identity in 591 aa overlap:

```

55  orf61a.pep  10 20 30 40 50 60
    MTVLKPSHWRVLAELADGLPQHVSQARMADMKPQQLNGFWQOMPAHIRGLLRQHDGYWR
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  orf61-1     MTVLKLSHWRVLAELADGLPQHVSQARMADMKPQQLNGFWQOMPAHIRGLLRQHDGYWR
    10 20 30 40 50 60

    70 80 90 100 110 120
  orf61a.pep  LVRPLAVFDAEGLRELTERS GFQTLKHECASSNDEILELARIAPDKAHKTICVTHLQSK
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  orf61-1     LVRPLAVFDAEGLRELTERS GFQTLKHECASSNDEILELARIAPDKAHKTICVTHLQSK
    70 80 90 100 110 120

65  130 140 150 160 170 180

```

-174-

	orf61a.pep	GRGRQGRKWSHRLGECCLMFSFGWVFDPRQYELGSLSPVAACRRALSRLGLKTQIKWPN	
	orf61-1	GRGRQGRKWSHRLGECCLMFSFGWVFDPRQYELGSLSPVAACRRALSRLGLDVQIKWPN	
5		130 140 150 160 170 180	
	orf61a.pep	DLVVGRDKLGGILIEVTRTGKTVAVVGIGINFLPKEVENAASVQSLFQTASRRGNADA	
10	orf61-1	DLVVGRDKLGGILIEVTRTGKTVAVVGIGINFLPKEVENAASVQSLFQTASRRGNADA	
		190 200 210 220 230 240	
	orf61a.pep	AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG	
15	orf61-1	AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG	
		250 260 270 280 290 300	
	orf61a.pep	QGVLEHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSEFLLLDGGNSRLKWAWVENGT	
20	orf61-1	QGVLEHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSEFLLLDGGNSRLKWAWVENGT	
		310 320 330 340 350 360	
	orf61a.pep	ATVGSAPYRDLSPGLAEWAEEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL	
25	orf61-1	ATVGSAPYRDLSPGLAEWAEEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL	
		370 380 390 400 410 420	
	orf61a.pep	GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDGHHYLGGTIMPGF	
30	orf61-1	GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDGHHYLGGTIMPGF	
		430 440 450 460 470 480	
	orf61a.pep	HLMKESLAVRTANLNRHAGKRYFPPTTTGNAVASGMMDAVCGSVMMHGRLEKKTGAGKP	
35	orf61-1	HLMKESLAVRTANLNRHAGKRYFPPTTTGNAVASGMMDAVCGSVMMHGRLEKKTGAGKP	
		490 500 510 520 530 540	
	orf61a.pep	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNIAAEGGESEHTX	
40	orf61-1	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNIMIAAEGGREYEHIX	
		550 560 570 580 590	

Homology with a predicted ORF from *N.gonorrhoeae*ORF61 shows 94.2% identity over a 189aa overlap with a predicted ORF (ORF61.ng) from *N.*50 *gonorrhoeae*:

	orf61.pep	EISLRSDXRPVSVXKRRDSEFLLLDGGNS	30
	orf61ng	TVCEGTVKGVDRGVLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPDSEFLLLEGGNS	211
55	orf61.pep	RLKWAWVENGTFATVGSAPYRDLSPGLAEWAEEKADGNVRIVGCAVCGEFKKAQVQEQLAR	90
	orf61ng	RLKWAWVENGTFATVGSAPYRDLSPGLAEWAEEKADGNVRIVGCAVCGESKKAQVQEQLAR	271
60	orf61.pep	KIEWLPSSAQAXGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTD	150
	orf61ng	KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTD	331
	orf61.pep	GHYLGXGTIMPGFHLMKESLAVRTANLNRHAGKRYFPPT	189
65	orf61ng	GHYLG-GTIMPGFHLMKESLAVRTANLNRPAKRYFPPTTTGNAVASGMMDAVCGSIMM	390

An ORF61ng nucleotide sequence <SEQ ID 237> was predicted to encode a protein having amino acid sequence <SEQ ID 238>:

```

1  MFSFGWAFDR PQYELGSLSP VAALACRRAL GCLGLETQIK WPNDLVVGRD
5  51  KLGGILITV RAGGKTVAVV GIGINFVLPK EVENAASVQS LFQTASRRGN
101 ADAAVLLET LLAELGAVLEQ YAEFGFAPFL NEYETANRDH GKAVLLLRDG
151 ETVCEGTVKG VDGRGVLHLE TAEGEQTVVS GEISLRPDNR SVSVPKRPDS
201 ERFLLEGGN SRLKWAVVEN GTFATVGSAP YRDLSP LGAE WAEKADGNVR
251 IVGCAVCGES KKAQVKEQLA RKIEWLPSSA QALGIRNHYR HPEEHGSDRW
301 FNALGSRRFS RNACVVVSCG TAVTVDALTD DGHYLG GTIM PGFHLMKESL
10 351 AVRTANLNR AGKRYPFPTT TGNVAVSGMM DAVCGSIMMM HGRLKEKNGA
401 GKPVDVIITG GGAAKVAEAL PPAFLAENTV RVADNLVIHG LLNLIAAEGG
451 ESEHA*

```

Further analysis revealed the complete gonococcal DNA sequence <SEQ ID 239> to be:

```

1  ATGACGGTTT TGAAGCCTTC GCATTGGCGG GTGTTGGCGG AGCTTGCCGA
15 51  CGGTTTGCCG CAACACGTAT CGCAATTGGC GCGTGAGGCG GACATGAAGC
101 CGCAGCAGCT CAACGGTTT TGGCAGCAGA TGCCGCGCGA TATACGCGGG
151 CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CCTTGCGGGT
201 TTTCGATGCC GAAGGTTTGC GCGATCTGGG GGAAAGGTCG GGTTTTCAGA
251 CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
20 301 GCGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGCG TGACCCACCT
351 GCAAAGTAAG GGCAGGGGCG GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
401 GCGAGTGCCT GATGTTCACT TCCGGCTGGG CGTTTGACCG GCCGCAGTAT
451 GAGTTGGGTT CGCTGTCGCC TGTTGCGGCA CTTGCGTGCC GGCGCGCTTT
501 GGGGTGTTTG GGTGTTGAAA CGCAAATCAA GTGGCCAAAC GATTTGGTCG
25 551 TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACAGT CAGGCGCGGC
601 GGTAAAACGG TTGCCGTGCT CGGTATCGGC ATCAATTTCG TGCTGCCCCA
651 GGAAGTGGAA AACGCCGCTT CCGTGCAGTC GCTGTTTCAG ACGGCATCGC
701 GCGCGGGCAA TGCCGATGCC GCCGTATTGC TGGAACATT GCTTGCGGAA
751 CTGGGCGCGG TGTTGGAACA ATATGCGGAA GAAGGGTTCG CGCCATTTTT
30 801 AAATGAGTAT GAAACGGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
851 TGCGCGACGG CGAAACCGTG TGCGAAGGCA CGGTAAAGG CGTGACGCGA
901 CGAGGCGTTC TGCATTGGA AACGGCAgaa ggcgaACAGa cgtcgtcag
951 cggcgaatC AGcctGCggc ccgacaacAG GTCGGtttcc gtgcgaagc
1001 ggccggatTC GgaacgtTTT tTGctgttgg aaggcgggaa cagccgGCTC
35 1051 AAGTGGGCGT GggtggAAaA cggcacgttc gcaaccgtgg gcagcgcgCc
1101 gtaCCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGGCGGATG
1151 GAAATGTCCG CATCGTCCGT TGCGCCGTGT GCGGAGAATC CAAAAGGCA
1201 CAAGTGAAGG AACAGCTCGC CCGAAAATC GAGTGGCTGC CGTCTCCGC
1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCGAA GAACACGGTT
40 1301 CCGACCGTTG GTTCAACGCC TTGGGCAGCC GCCGCTTCAG CCGCAACGCC
1351 TCGCTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTACCGA
1401 TGACGGACAT TATCTCGGCG GAACCATCAT GCGGCGCTTC CACCTGATGA
1451 AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGCCC CGCCGGCAAA
1501 CGTTACCCTT TCCCGACCAC AACGGGCAAC GCCGTCGAA GCGGCATGAT
45 1551 GGACGCGGTT TGCGGCTCGA TAATGATGAT GCACGGCCGT TTGAAAGAAA
1601 AAAACGGCGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG
1651 GCGAAAGTCG CCGAAGCCCT GCCGCCTGCA TTTTGGCGG AAAATACCGT
1701 GCGCGTGGCG GACAACCTCG TCATCCACGG GCTGCTGAAC CTGATTGCCG
1751 CCGAAGGCGG GGAATCGGAA CACGCTTAA

```

50 This corresponds to the amino acid sequence <SEQ ID 240; ORF61ng-1>:

```

1  MTVLKPSHR VLAELADGLP QHVSQ LAREA DMKPQQLNGF WQQMPAHIRG
51  LLRQHDGYWR LVRPLAVFDA EGLRDLGERS GFQTALKHEC ASSNDEILEL
101 ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWAFDRPQY
55 151 ELGSLSPVAA LACRRALGCL GLETQIKWPN DLVVGRDKLG GILITVVRAG
201 GKTVAVVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLAE
251 LGAVLEQYAE EGFAPFLNEY ETANRDHGKA VLLLRDGETV CEGTVKGVVDG
301 RGVLHLETAE GEQTVVSGEI SLRPNDRSVS VPKRPDSERF LLLEGNSRL
351 KWAVWENGTF ATVGSAPYRD LSPLGAEWAE KADGNVRIVG CAVCGESKKA
401 QVKEQLARKI EWLPSAQAL GIRNHYRHPE EHGS DRWFNA LGSRRFSRNA
60 451 CVVVSCTAV TVDALTDGHH YLGGTIMPGF HLMKESLAVR TANLNRPAKG
501 RYFPFTTTGN AVASGMMDAV CGSIMMMHGR LKEKNGAGKP VDVIITGGGA
551 AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HA*

```

ORF61ng-1 and ORF61-1 show 93.9% identity in 591 aa overlap:

	orf61ng-1.pep	MTVLKPSHWRVLAELADGLPQHVSQALAREADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR	60
	orf61-1	MTVLKLSHWRVLAELADGLPQHVSQALARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR	60
5	orf61ng-1.pep	LVRPLAVFDAEGLRDLGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK	120
	orf61-1	LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK	120
10	orf61ng-1.pep	GRGRQGRKWSHRLGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPN	180
	orf61-1	GRGRQGRKWSHRLGECLMFSFGWVDFRPQYELGSLSPVAACRRALSRLGLDVQIKWPN	180
15	orf61ng-1.pep	DLVVGRDKLGGILIIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	240
	orf61-1	DLVVGRDKLGGILIIETVRTGCKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	240
	orf61ng-1.pep	AVLLETLLAELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVVDG	300
20	orf61-1	AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVVDG	300
	orf61ng-1.pep	RGVLHLETAEGEQTVVSGEISLRPDNRSVSPKRPDSEFLLLEGNSRLKWAVVENGTFF	360
	orf61-1	QGVLHLETAEGKQTVVSGEISLRSDRPVSPKRRDSEFLLLDGNSRLKWAVVENGTFF	360
25	orf61ng-1.pep	ATVGSAPYRDLSPGLAEWAEEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQAL	420
	orf61-1	ATVGSAPYRDLSPGLAEWAEEKADGNVRIVGCAVCGEFKKAQVQEQQLARKIEWLPSSAQAL	420
30	orf61ng-1.pep	GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDGHHYLGGTIMPGF	480
	orf61-1	GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDGHHYLGGTIMPGF	480
	orf61ng-1.pep	HLMKESLAVRTANLNRPAKRYFPFTTTGNAVASGMMDAVCGSIMMMHGRLEKNGAGKP	540
35	orf61-1	HLMKESLAVRTANLNRHAGRYFPFTTTGNAVASGMMDAVCGSVMMMHGRLEKKTGAGKP	540
	orf61ng-1.pep	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNIAAEGGESEHAX	593
40	orf61-1	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNIMIAAEGREYEHIX	593

Based on this analysis, including the homology with the baf protein of *B.pertussis* and the presence of a putative prokaryotic membrane lipoprotein lipid attachment site, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 45 Example 29

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 241>:

	1	ATGTTTACC	AAATCCTGC	CCTGATTATC	TGGAGCAGCT	CGTTTATTGC
	51	CGCCAAATAT	GTCTATGGCG	GCATCGATCC	CGCATTGATG	GTCGGCGTGC
50	101	GCCTGCTAAT	TGCCGCGCTG	CCTGCACTGC	CCGCTGCGG	CCGTCATGTC
	151	GGCAAGATTC	CGCGTGAGGA	ATGGAAGCCG	TTGCTGATTG	TGTCGTTTCGT
	201	CAACTATGTG	CTGACCCTGC	TGCTTCAGTT	TGTCGGGTTG	AAATACACTT
	251	CCGCCGCCAG	CGCATCGGTC	ATTGTCGGAC	TCGAGCCGCT	GCTGATGGTG
	301	TTTGTCGGAC	ACTTTTCTT	CAACGACAAA	GCGCGTGCCT	ACCACTGGAT
55	351	ATGCGGCGCG	GCGGCATTTC	CCGGTGTGCG	GCTGCTGATG	GCGGGCGGTG
	401	CGGaAGAGGG	CGGCgaAGTC	GGCTGGTTCG	GCTGCCTGCT	GGTGTGTGTTG
	451	GCGGGCGCGG	GCTTTTGTGC	CGCTATGCGT	CCGACGCAAA	GGCTGATTGC
	501	ACGCATCGGC	GCACCGGCAT	TCACATCTGT	TTCCATTGCC	GCCGCATCGT
	551	TGATGTGCCT	GCCGTTTTTCG	CTTGCTTTGG	CGCAAAGTTA	TACCGTGGAC
60	601	TGGAGCGTCG	GGATGCTATT	GTGCTGCTG	TATTTGGGTT	TGGGGTGC..

This corresponds to the amino acid sequence <SEQ ID 242; ORF62>:

1 MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLIIAAL PALPACRRHV  
 51 GKIPREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV  
 101 FVGHHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL  
 151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD  
 5 201 WSVGMVLSLL YLGLGC..

Further work revealed the complete nucleotide sequence <SEQ ID 243>:

1 ATGTTTTACC AAATCCTTGC CCTGATTATC TGGAGCAGCT CGTTTTATTGC  
 51 CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC  
 101 GCCTGCTAAT TGCCGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC  
 151 GGCAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTCTG  
 201 CAACTATGTG CTGACCCTGC TGCTTCAGTT TGTCGGGTG AAATACACTT  
 251 CCGCCGCCAG CGCATCGGTC ATTGTCGGAC TCGAGCCGCT GCTGATGGTG  
 301 TTTGTCGGAC ACTTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT  
 351 ATGCGGCGCG GCGGCATTG CCGGTGTCG GCTGCTGATG GCGGGCGGTG  
 15 401 CGGAAGAGGG CGGCGAAGTC GGCTGGTTCG GCTGCCTGCT GGTGTTGTTG  
 451 GCGGGCGCGG GCTTTTGTGC CGCTATGCGT CCGACGCAA GGCTGATTGC  
 501 ACGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT  
 551 TGATGTGCCT GCCGTTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGGAC  
 601 TGGAGCGTCG GGATGGTATT GTCGCTGCTG TATTTGGGTT TGGGGTGGCG  
 20 651 CTGGTACGCC TATTGGCTGT GGAACAAGGG GATGAGCCGT GTTCCTGCCA  
 701 ATGTTTCGGG ACTGTTGATT TCGCTCGAAC CCGTCGTCGG CGTGCTGCTG  
 751 GCGGTTTTGA TTTTGGGCGA ACACCTGTCG CCCGTGTCCG CCTTGGGCGT  
 801 GTTGTGCTC ATCGCCGCCA CCTTGGTTGC CGGCCGGCTG TCGCATCAAA  
 851 AATAA

25 This corresponds to the amino acid sequence <SEQ ID 244; ORF62-1>:

1 MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLIIAAL PALPACRRHV  
 51 GKIPREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV  
 101 FVGHHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL  
 151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD  
 30 201 WSVGMVLSLL YLGLGCGWYA YWLWNKMSR VPANVSGLLI SLEPVVGVL  
 251 AVLILGEHLS PVSALGVFV IAATLVAGRL SHQK\*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical transmembrane protein HI0976 of *H. influenzae* (accession number Q57147)

ORF62 and HI0976 show 50% aa identity in 114aa overlap:

35 Orf62 1 MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHV  
 M YQILAL+IWSSS I K Y +DP L+V VR R KI + K  
 HI0976 1 MLYQILALLIWSSSLIVGKLTYSMMDPVLVQVRLIIAMIIVMPLFLRRWKKIDKPMRKQ 60  
 40 Orf62 61 LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFGHHFFNDKARAY 114  
 L ++F NY LLQF+GLKYTSA+SA ++GLEPLL+VFVGHFFF K +  
 HI0976 61 LWWLAFFNYTAVFLLQFIGLKYTSAASAVTMIGLEPLLVVFGHHFFKTKQNGF 114

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF62 shows 99.5% identity over a 216aa overlap with an ORF (ORF62a) from strain A of *N.*

45 *meningitidis*:

10 20 30 40 50 60  
 orf62.pep MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHV  
 50 orf62a MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHV  
 10 20 30 40 50 60  
 70 80 90 100 110 120  
 orf62.pep LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFGHHFFNDKARAYHWICGA  
 55 orf62a LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFGHHFFNDKARAYHWICGA  
 70 80 90 100 110 120  
 130 140 150 160 170 180  
 orf62.pep AAFAGVALLMAGGAEEGGEVGVFGCLLVLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA

-178-

```

      |||
orf62a  AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA
      130      140      150      160      170      180
5
      190      200      210
orf62.pep  AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC
      |||
orf62a  AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVGCSWYAYWLWNKGMSRVPANVSGLLI
      190      200      210      220      230      240
10
orf62a  SLEPVVGVLAVLILGEHLSPVSVLGVFVVAATLVAGRLSHQKX
      250      260      270      280

```

The complete length ORF62a nucleotide sequence <SEQ ID 245> is:

```

15      1  ATGTTTACC  AAATCCTTGC  CCTGATTATC  TGGAGCAGCT  CGTTTATTGC
      51  CGCCAAATAT  GTCTATGGCG  GCATCGATCC  CGCATTGATG  GTCGGCGTGC
      101  GCCTGCTGAT  TGCTGCGCTG  CCTGCACTGC  CCGCCTGCCG  CCGTCATGTC
      151  GGCAAGATTC  CGCGTGAGGA  ATGGAAGCCG  TTGCTGATTG  TGTCGTTTCGT
      201  CAACTATGTG  CTGACCCTGC  TACTTCAGTT  TGTCGGGTG  AAATACACTT
      251  CCGCCGCCAG  CGCATCGGTC  ATTGTCGGAC  TCGAGCCACT  GCTGATGGTG
20      301  TTTGTCGGAC  ACTTTTCTT  CAACGACAAA  GCGCGTGCCT  ACCACTGGAT
      351  ATGCGGCGCG  GCGGCATTTC  CCGGTGTCGC  GCTGCTGATG  GCGGGCGGTG
      401  CGGAAGAGGG  CGGCGAAGTC  GGCTGGTTCG  GCTGCCTGCT  GGTGTTGTTG
      451  GCGGGCGCGG  GCTTTTGTGC  CGCTATGCGT  CCGACGCAAA  GGCTGATTGC
      501  ACGCATCGGC  GCACCGGCAT  TCACATCTGT  TTCCATTGCC  GCCGCATCGT
25      551  TGATGTGCCT  GCCGTTTCG  CTGCTTTGG  CGCAAAGTTA  TACCGTGGAC
      601  TGGAGCGTCG  GAATGGTATT  GTCGCTGCTG  TATTTGGGCG  TGGGGTGCAG
      651  CTGGTACGCC  TATTGGCTGT  GGAACAAGGG  GATGAGCCGT  GTTCCTGCCA
      701  ACGTTTCGGG  ACTGTTGATT  TCGCTCGAAC  CCGTCGTCGG  CGTGCTGCTG
      751  GCGGTTTGA  TTTTGGGCGA  ACACCTGTCG  CCCGTGTCCG  TCTTGGGCGT
30      801  GTTTGTGTC  ATCGCCGCCA  CCTTGGTTGC  CGGCCGGCTG  TCGCATCAA
      851  AATAA

```

This encodes a protein having amino acid sequence <SEQ ID 246>:

```

35      1  MFYQILALII  WSSSFIAAKY  VYGGIDPALM  VGVRLLIAAL  PALPACRRHV
      51  GKIPIREWK  LLIVSFVNYV  LTLQLQFVGL  KYTSAASASV  IVGLEPLLMV
      101  FVGHHFFNDK  ARAYHWICGA  AAFAGVALLM  AGGAEEGGEV  GWFGCLLVLL
      151  AGAGFCAAMR  PTQRLIARIG  APAFTSVSIA  AASLMCLPFS  LALAQSYTVD
      201  WSVGMVLSLL  YLGVGCSWYA  YWLWNKGMSR  VPANVSGLLI  SLEPVVGVL
      251  AVLILGEHLS  PVSVLGVFVV  IATLVAGRL  SHQK*

```

ORF62a and ORF62-1 show 98.9% identity in 284 aa overlap:

```

40      orf62a.pep  MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHV 60
      orf62-1      MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHV 60
45      orf62a.pep  LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA 120
      orf62-1      LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA 120
      orf62a.pep  AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA 180
      orf62-1      AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA 180
50      orf62a.pep  AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVGCSWYAYWLWNKGMSRVPANVSGLLI 240
      orf62-1      AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANVSGLLI 240
55      orf62a.pep  SLEPVVGVLAVLILGEHLSPVSVLGVFVVAATLVAGRLSHQKX 285
      orf62-1      SLEPVVGVLAVLILGEHLSPVSVLGVFVVAATLVAGRLSHQKX 285
60

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF62 shows 99.5% identity over a 216aa overlap with a predicted ORF (ORF62.ng) from *N. gonorrhoeae*:



-179-

	orf62.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVKGKIPREEWKP	60
	orf62ng	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVKGKIPREEWKP	60
5	orf62.pep	LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMLVFVGHFFNDKARAYHWICGA	120
	orf62ng	LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMLVFVGHFFNDKARAYHWICGA	120
10	orf62.pep	AAFAGVALLMAGGAEEGGEVGVFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFSTVSIA	180
	orf62ng	AAFAGVALLMAGGAEEGGEVGVFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFSTVSIA	180
	orf62.pep	AASLMCLPFSLALAAQSYTVDWSVGMVLSLLYLGLGC	216
15	orf62ng	AASLMCLPFSLALAAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI	240

The complete length ORF62ng nucleotide sequence <SEQ ID 247> is:

	1	ATGTTTACC	AAATCCTTGC	CCTGATTATC	TGGGGCAGCT	CGTTTATTGC
	51	CGCCAAATAT	GTCTATGGCG	GCATCGATCC	CGCATTGATG	GTGCGCGTGC
20	101	GCCTGCTGAT	TGCCGCGCTG	CCTGCACTGC	CCGCCTGCCG	CCGTCATGTC
	151	GGCAAGATTC	CGCGTGAGGA	ATGGAAGCCG	TTGCTGATTG	TGTCGTTTCGT
	201	CAACTATGTG	CTGACCCCTG	TGCTTCAGTT	TGTCGGGTTG	AAATACACTT
	251	CCGCCGCCAG	CGCATCGGTC	ATTGTCGGAC	TCGAGCCGCT	GCTGATGGTG
	301	TTTGTGCGAC	ACTTTTTCTT	CAACGACAAA	GCGCGTGCCT	ACCACTGGAT
	351	ATGCGGCGCG	GCGGCATTTC	CCGGTGTCGC	GCTGCTGATG	GCGGGCGGTG
25	401	CGGAAGAGGG	CGGCGAAGTC	GGCTGGTTCG	GCTGCTGCTG	GGTGTGTTG
	451	GCGGGCGCGG	GCTTTTGTGC	CGCTATGCGT	CCGACGCAAA	GGCTGATTGC
	501	CCGCATCGGC	GCACCGGCAT	TCACATCTGT	TTCCATTGCC	GCCGCATCGT
	551	TGATGTGCCT	GCCGTTTTTC	CTTGCTTTGG	CGCAAAGTTA	TACCGTGGAC
	601	TGGAGCGTCG	GGATGGTATT	GTCGCTGTTG	TATTTGGGTT	TGGGGTGCCG
30	651	CTGGTACGCC	TATTGGCTGT	GGAACAAGGG	GATGAGCCGT	GTTCTGSCCA
	701	ACGCGTCGGG	ACTGTTGATT	TCGCTCGAAC	CCGTCGTCGG	CGTGCTGTTG
	751	GCGGTTTTGA	TTTTGGGCGA	ACATTTATCG	CCCGTGTCGG	CCTTGGGCGT
	801	GTTTGTGCTC	ATCGCCGCCA	CTTTCGCCGC	CGGCCGGCTG	TCGCGCAGGG
	851	ACGCGCAAAA	CGGCAATGCC	GTCTGA		

35 This encodes a protein having amino acid sequence <SEQ ID 248>:

	1	<u>MFYQILALII</u>	<u>WGSSFIAAKY</u>	<u>VYGGIDPALM</u>	<u>VGVRLLIAAL</u>	<u>PALPACRRHV</u>
	51	<u>GKIPREEWKP</u>	<u>LLIVSFVNYV</u>	<u>LTLLLQFVGL</u>	<u>KYTSAAASV</u>	<u>IVGLEPLLMLV</u>
	101	<u>FVGHFFNDK</u>	<u>ARAYHWICGA</u>	<u>AAFAGVALLM</u>	<u>AGGAEEGGEV</u>	<u>GWFGCLLVLL</u>
	151	<u>AGAGFCAAMR</u>	<u>PTQRLIARIG</u>	<u>APAFSTVSIA</u>	<u>AASLMCLPFS</u>	<u>LALAAQSYTVD</u>
40	201	<u>WSVGMVLSLL</u>	<u>YLGLGCGWYA</u>	<u>YWLWNKGMSR</u>	<u>VPANASGLLI</u>	<u>SLEPVVGVLL</u>
	251	<u>AVLILGEHLS</u>	<u>PVSALGVFVV</u>	<u>IAATFAAGRL</u>	<u>SRRDAQNGNA</u>	<u>V*</u>

ORF62ng and ORF62-1 show 97.9% identity in 283 aa overlap:

		10	20	30	40	50	60
45	orf62ng.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVKGKIPREEWKP					
	orf62-1	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVKGKIPREEWKP					
		10	20	30	40	50	60
50	orf62ng.pep	LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMLVFVGHFFNDKARAYHWICGA					
	orf62-1	LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMLVFVGHFFNDKARAYHWICGA					
		70	80	90	100	110	120
55	orf62ng.pep	AAFAGVALLMAGGAEEGGEVGVFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFSTVSIA					
	orf62-1	AAFAGVALLMAGGAEEGGEVGVFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFSTVSIA					
60		130	140	150	160	170	180
	orf62ng.pep	AASLMCLPFSLALAAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI					
	orf62-1	AASLMCLPFSLALAAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI					
65		190	200	210	220	230	240
	orf62ng.pep	AASLMCLPFSLALAAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI					
	orf62-1	AASLMCLPFSLALAAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI					

-180-

```

                250      260      270      280      290
orf62ng.pep    SLEPVVGVLAVLILGEHLSVPSALGVFVIAATFAAGRLSRRDAQNGNAVX
5             orf62-1    SLEPVVGVLAVLILGEHLSVPSALGVFVIAATLVAGRLSHQKX
                250      260      270      280

```

Furthermore, ORF62ng shows significant homology to a hypothetical *H. influenzae* protein:

```

10  sp|Q57147|Y976_HAEIN HYPOTHETICAL PROTEIN HI0976 >gi|1074589|pir||B64163
    hypothetical protein HI0976 - Haemophilus influenzae (strain Rd KW20)
    >gi|1574004 (U32778) hypothetical [Haemophilus influenzae] Length = 128
    Score = 106 bits (262), Expect = 2e-22
    Identities = 56/114 (49%), Positives = 68/114 (59%)

15  Query: 1   MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRRXXXXXXXXXXXXCRRHVGVKIPREEWKP 60
      M YQILAL+IW SS I K Y +DP L+V VR R KI + K
      Sbjct: 1   MLYQILALLIWSSSLIVGKLTYSMMDPVLVVQVRLIAMIIVMPLFLRRWKIDKPMRKQ 60

20  Query: 61   LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMEVFGHFFFNDKARAY 114
      L ++F NY LLQF+GLKYTSA+SA ++GLEPLL+VFGHFFF K +
      Sbjct: 61   LWLWLAFFNYTAVFLLQFIGLKYTSASSAVTMIGLEPLLVEVFGHFFFKTKQNGF 114

```

Based on this analysis, including the homology with the transmembrane protein of *H. influenzae* and the putative leader sequence and several transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 30

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 249>:

```

30  1   ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCmGwms TCCTGkkGTA
    51   sGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
    101  GGTGGATTGT TGCGTTCAGC GCAATGCTGC TGCTGGTGTT GTCCGCCGTT
    151  TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT
    201  CGGTTCGCTa srTyGCCAAA gsGCCTgkks TGGG.ATGTT TACGCTGGTT
    251  GCGGkACTGC CCGGCGTGTT TCTGTTTCGGC TTTCCCGCAC AGTTCATCAA
35  301  CGGCACGATT AATTCGTGGT TCGGCAACGA TACCCACGAG GCGCTTGAAC
    351  GCAGCCTCAA TTTGAGCAAG TCCGCATTGA ATTTGGCGGC AGACAACGCC
    401  CTCGGCAACG CCGTCCCCGT GCAGATAGAC CTCATCGCGC CGGCTTCCCT
    451  GCCCGGGGAT ATGGGCAGGG TGCTGGAACA TTACGCCGGC AGCGGTTTTG
    501  CCGAGCTTGC CCTGTACAAY ksCGCAAGCG GCAAAATCGA AAAAAAGCATC
    551  AACCCGCACA AGCTCGATCA GCCGTTTCCA GGTAAGGCGC GTTGGGAaAa
40  601  AATCCaACGG GCGGGTTCGG TCAGGGATTG GGAAAGCATA GCGGCGGTAT
    651  TGTaCGCGCA GGGCTGGCTG TCGGCGGGTA CGCACwACGG GCGCGATTAC
    701  GCCTTGTTTT TCCGTCAGCC GGTTCCCAAA GGCGTGGCAG AGGATGCCGT
    751  yTTAATCGAA AAGGCAAGGG CGAAATATGC TGAGTTGAGT TACAGCAAAA
45  801  AAGGTTTGCA GACCTTTTTT CTGGCAACCC TGCTGATTGC CTCGCTGCTG
    851  TCGATTTTTT TTGCACTGGT CATGGCACTG TATTTGCCCC GCCGTTTCGT
    901  CGAACCCGTC CTATCGCTTG CCGAGGGGGC GAAGGCGGTG GCGCAAGGCG
    951  ATTTAGCCA GACGCGCCCC GTGTTGCGCA ACGACGAGTT CGGACGCTTG
50  1001 ACCArGTTGT TCAACCATAT GACCGAGCAG CTTTCCATCG CCAAAGATGC
    1051 AGACGAGCGC AACC GCCGGC GCGAGGAAGC CGCCAGGCAT TATCTTGAAT
    1101 GCGTGTTGGA GGGGCTGACC ACGGGCGTGG TGGTGTTTGA CGAACAAGGC
    1151 TGTCTGAAAA CCTTCAACAA AGCGGCGGGT ACC..

```

This corresponds to the amino acid sequence <SEQ ID 250; ORF64>:

```

55  1   MRRFLPIAAI CAXXLXXGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
    51   LARYVILLK DRRDGVFGSX XAKXPXXXMF TLVAXLPGVF LFGFPAQFIN
    101  GTINSWFGND THEALERSLN LSKSALNLAA DNALGNAVVP QIDLIGAASL
    151  PGDMGRVLEH YAGSGFAQLA LYNXASGKIE KSINPHKLDQ PFPKGARWEK
    201  IQRAGSVRDL ESIGGVLYAQ GWLSAGTHXG RDYALFFRQP VPKGVAEDAV
    251  LIEKARAKYA ELSYSKKGLO TFFLATLLIA SLLSIFLALV MALYFARRFV

```

301 EPVLSLAEGA KAVAQGDFSQ TRPVLNRNDEF GRLTXLFNHM TEQLSIAKDA  
 351 DERNRRREEA ARHYLECVLE GLTTGVVVFDE EQGCLKTFNK AAGT..

Further work revealed the complete nucleotide sequence <SEQ ID 251>:

5 1 ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCCGTCG TCCTGTTGTA  
 51 CGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT  
 101 GGTGGATTGT TGCGTTCAGC GCAATGCTGC TGCTGGTGTG GTCCGCCGTT  
 151 TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT  
 201 CGGTTCGCAG ATTGCCAAAC GCCTTCTCTG GATGTTTACG CTGGTTGCCG  
 251 TACTGCCCGG CGTGTTCCTG TTCGCGGTTT CCGCACAGTT CATCAACGGC  
 10 301 ACGATTAATT CGTGGTTCGG CAACGATACC CACGAGGCGC TTGAACGCAG  
 351 CCTCAATTTG AGCAAGTCCG CATTGAATTT GGCGGCAGAC AACGCCCTCG  
 401 GCAACGCCGT CCCCCTGCAG ATAGACCTCA TCGGCGCGGC TTCCTGCCCC  
 451 GGGGATATGG GCAGGGTGCT GGAACATTAC GCCGGCAGCG GTTTTGCCCA  
 501 GCTTGCCCTG TACAATGCCG CAAGCGGCAA AATCGAAAAA AGCATCAACC  
 15 551 CGCACAAAGCT CGATCAGCCG TTTCCAGGTA AGGCGCGTTG GGAAAAAATC  
 601 CAACGGGCGG GTTCGGTCAG GGATTGGAA AGCATAGGCG GCGTATTGTA  
 651 CGCGCAGGGC TGGCTGTCGG CGGGTACGCA CAACGGGCGC GATTACGCCT  
 701 TGTTTTTCCG TCAGCCGGTT CCCAAAGGCG TGGCAGAGGA TGCCGTCTTA  
 751 ATCGAAAAGG CAAGGGCGAA ATATGCTGAG TTGAGTTACA GCAAAAAAGG  
 20 801 TTTGCAGACC TTTTCTCTGG CAACCTGCT GATTGCCTCG CTGCTGTCGA  
 851 TTTTCTTTCG ACTGGTCATG GCACTGTATT TCGCCGCGCG TTTCGTCGAA  
 901 CCCGTCCATAT CGCTTGCCGA GGGGCGGAAG GCGGTGGCGC AAGGCGATTT  
 951 CAGCCAGACG CGCCCCGTGT TGCGCAACGA CGAGTTCGGA CGCTTGACCA  
 1001 AGTTGTTCAA CCACATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC  
 25 1051 GAGCGCAACC GCCGGCGCGA GGAAGCCGCC AGGCATTATC TTGAATGCGT  
 1101 GTTGGAGGGG CTGACCACGG GCGTGGTGGT GTTGACGAA CAAGGCTGTC  
 1151 TGAAAACCTT CAACAAAGCG GCGGAACAGA TTTTGGGGAT GCCGCTTACC  
 1201 CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA  
 1251 GTCCCTGCTT GCCGAAGTGT TTGCCGCCAT CGGCGCGGCG GCAGGTACGG  
 30 1301 ACAAACCGGT CCATGTGAAA TATGCCGCGC CGGACGATGC CAAAATCCTG  
 1351 CTGGGCAAGG CAACCGTCTT GCCGAAGAC AACGGCAACG GCGTGGTAAT  
 1401 GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAAA GAAGCCGCGT  
 1451 GGGGCGAAGT GGCGAAGCGG CTGGCACACG AAATCCGCAA TCCGCTCAGC  
 35 1501 CCCATCCAGC TTTCCGCGA ACAGGCTGGC TGGAATTTGG GCGGGAAGCT  
 1551 GGATGAGCAG GATGCGCAAA TCCTGACGCG TTCGACCGAC ACCATCGTCA  
 1601 AACAGGTGGC GGCATTGAAG GAAATGGTCG AAGCATTCCG CAATTATGCG  
 1651 CGTTCCCCTT CGCTCAAATT GGAAATCAG GATTTGAACG CCTTAATCGG  
 1701 CGATGTGTTG GCATTGTATG AAGCCGGTCC GTGCCGTTT GCGGCGGAGC  
 40 1751 TTGCCGCGCA ACCGCTGACG GTGGCGGCGG ATACGACCGC CATGCGGCAG  
 1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA  
 1851 TGTGCCCAGG GTCAGGTAA AATCGGAAAC AGGGCAGGAC GGTCCGATTG  
 1901 TCCTGACGGT TTGCGACAAC GGCAAAGGGT TCGGCAGGGA AATGCTGCAC  
 1951 AACGCCTTCG AGCCGTATGT AACGGACAAA CCGGCGGGAA CGGGATTGGG  
 45 2001 TCTGCCTGTG GTGAAAAAAA TCATTGAAGA ACAGGCGGCG CGCATCAGCC  
 2051 TGAGCAATCA GGATGCGGGT GGCGCGTGTG TCAGAATCAT CTTGCCAAAA  
 2101 ACGGTAAAAA CTTATGCGTA G

This corresponds to the amino acid sequence <SEQ ID 252; ORF64-1>:

50 1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV  
 51 LARYVILLK DRRDGVFGSQ IAKRLSGMFT LVAVLPGVFL FGVSQFING  
 101 TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNAVVPQ IDLIGAASLP  
 151 GDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI  
 201 QRAGSVRDLE SIGGVLYAQG WLSAGTHNGR DYALFFRQPV PKGVAEDAVL  
 251 IEKARAKYAE LSYSKKGLQT FFLATLLIAS LLSIFLALVM ALYFARRFVE  
 301 PVLSLAEGAK AVAQGDFSQT RPVLNRNDEF RLTKLFNHMT EQLSIAKEAD  
 351 ERNRRREEAA RHYLECVLEG LTTGVVVFDE QGCLKTFNKA AEQILGMPLT  
 401 PLWGSSRHGW HGVSAAQSSL AEVFAAIGAA AGTDKPVHVK YAAPDDAKIL  
 451 LGKATVLPED NGNGVVMVID DITVLIHAQK EAAWGEVAKR LAHEIRNPLT  
 501 PIQLSAERLA WKLGGKLDQV DAQILTRSTD TIVKQVAALK EMVEAFRNYA  
 55 551 RPSCLKLENQ DLNALIGDVL ALYEAGPCRF AAELAGEPLT VAADTTAMRQ  
 60 601 VLHNIFKNAA EAAEEADVPE VRVKSETGQD GRIVLTVCDN GKGFGREMLH  
 651 NAFEPYVTDK PAGTGLGLPV VKKIIEEHGG RISLSNQDAG GACVRIILPK  
 701 TVKTYA\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF64 shows 92.6% identity over a 392aa overlap with an ORF (ORF64a) from strain A of *N.meningitidis*:

5	orf64.pep	MRRFLPIAAICAXLXXGLTAATGSTSSLADYFWWIVAFSAML	10	20	30	40	50	60
	orf64a	MRRFLPIAAICAVLLYGLTAATGSTSSLADYFWWIVAFSAML	10	20	30	40	50	60
10	orf64.pep	DRRDGVFGSXXAKXPXXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN	70	80	90	100	110	120
	orf64a	DRRDGVFGSQAIR-LSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLN	70	80	90	100	110	
15	orf64.pep	LSKSALNLAADNALGNAVVPQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNXASGKIE	130	140	150	160	170	180
	orf64a	LSKSALNLAADNALGNAIPVQIDXIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIE	120	130	140	150	160	170
20	orf64.pep	KSINPHKLDQFPFGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP	190	200	210	220	230	240
	orf64a	KSINPHKLDQFPFGKARWEKIQAGSVRDXESIGGVLYAXGWLSAXTHNGRDYALFFRQP	180	190	200	210	220	230
30	orf64.pep	VPKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV	250	260	270	280	290	300
	orf64a	VPKGVAEDAVLIEKARAXXXLSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV	240	250	260	270	280	290
35	orf64.pep	EPVLSLAEGAKAVAQGDIFSQTRPVLNRNDEFGRLTXLFNHMTQLSIAKDADERNRRREEA	310	320	330	340	350	360
	orf64a	EPVLSLAEGAKAVAQGDIFSQTRPVLNRNDEFGRLTXLFNHMTQLSIAKEADERNRRREEA	300	310	320	330	340	350
40	orf64.pep	ARHYLECVLEGLTTGVVVFDEQGLKTFNKAAGT	370	380	390			
	orf64a	ARHYLECVLEGLTTGVVVFDEQGLKTFNKAEEQILGMPLTPLWGSSRHGWHGVSAAQSL	360	370	380	390	400	410
45	orf64a	LAEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNXNGVVMVIDDITVLIHAQ	420	430	440	450	460	470

The complete length ORF64a nucleotide sequence <SEQ ID 253> is:

50	1	ATGCGCCGTT	TTCTACCGAT	CGCAGCCATA	TGCGCCGTCG	TCCTGTTGTA
	51	CGGACTGACG	GCGGCAACCG	GCAGCACCAG	TTCGCTGGCG	GATTATTTCT
	101	GGTGGATTGT	TGCGTTCAGC	GCAATGCTGC	TGCTGGTGTT	GTCCGCCGTT
	151	TTGGCACGTT	ATGTCATATT	GCTGTGAAA	GACAGGCGCG	ACGGCGTATT
55	201	CGGTTCCGAG	ATTGCCAAC	GCCTTCCGG	GATGTTTACG	CTGGTTGCCG
	251	TACTGCCCGG	CGTGTTCCTG	TTCGGCGTTT	CCGCACAGTT	TATCAACGGC
	301	ACGATTAATT	CGTGGTTCGG	CAACGATACC	CACGAGGCGC	TTGAACGCAG
	351	CCTCAATTG	AGCAAGTCCG	CATTGAATCT	GGCGGCAGAC	AACGCCCTTG
60	401	GCAACGCCAT	CCCCGTGCAG	ATAGACNTCA	TCCGCGCGGC	TCCCTGCC
	451	NGGGATATGG	GCAGGGTGCT	GGAACATTAC	GCCGGCAGCG	GTTTGGCCCA
	501	GCTTGCCCTG	TACAATGCCG	CAAGCGGCAA	AATCGAAAAA	AGCATCAACC
	551	CGCACAAGCT	CGATCAGCCG	TTTCCAGGTA	AGGCGCGTTG	GGAAAAAATC
65	601	CAACAGGCGG	GTTCGGTCAG	GGATNNGGAA	AGCATAGGCG	GCGTATTGTA
	651	CGCGCANGGC	TGGCTGTCGG	CAGNNACGCA	CAACGGGCGC	GATTACGCCT
	701	TGTTTTTCCG	TCAGCCGGTT	CCCAAAGGCG	TGGCAGAGGA	TGCCGTCTTA
	751	ATCGAAAAGG	CAAGGGCGNA	ANANNNTNAG	TTGAGTTACA	GCAAAAAGG
	801	TTTGCAGACC	TTTTTCTTNG	CAACCCTGCT	GATTGCCTCN	CTGCTGCGA
	851	TTTTTCTTGC	ACTGGTCATG	GCACTGTATT	TCGCCCGCCG	TTTCGTGAA

-183-

5  
 10  
 15  
 20  
 25

```

901 CCCGTCCTAT CGCTTGCCGA GGGGGCGAAG GCGGTGCGC AAGGCGATTT
951 CAGCCAGACG CGCCCCGTGT TGCGCAACGA CGAGTTCGGA CGCTTGACCA
1001 AGTTGTTCAA CCACATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC
1051 GAGCGCAACC GCCGGCGCGA GGAAGCCGCC AGACATTATC TCGAATGCGT
1101 GTTGGAGGGG CTGACCACGG GCGTGGTGGT GTTTGACGAA CAAGGTGTC
1151 TGAAAACCTT CAACAAAGCG GCGGAACAGA TTTTGGGGAT GCCGCTTACC
1201 CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
1251 GTCCCTGCTT GCCGAAGTGT TTGCCGCCAT CGGCGCGGCG GCAGGTACGG
1301 ACAAACCGGT CCATGTGAAA TATGCCGCGC CGGACGATGC CAAAATCCTG
1351 CTGGGCAAGG CAACCGTCCT GCCCGAAGAC AACNGCAACG GCGTGGTAAT
1401 GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAAA GAAGCCGCGT
1451 GGGGCGAAGT GGCAAAACGG CTGGCACACG AAATCCGCAA TCCGCTCACG
1501 CCCATCCAGC TTTCTGCCGA ACGGCTGGCG TGGAAATTGG GCGGGAAGCT
1551 GGACGAGCAN GACGCGCAAA TCCTGACACG TTCGACCGAC ACCATCATCA
1601 AACAAGTGGC GGCATTAAAA GAAATGGTCG AGGCATTCCG CAATTACNCG
1651 CGTTCCCTT CGNCTCAATT GGAAATCAG GATTTGAACG CCTTAATCGG
1701 CGATGTGTTG GCATTGTACG AAGCTGGTCC GTGCCGTTT GCGGCGGAAC
1751 TTGCCGGCGA ACCGCTGATG ATGGCGGCGG ATACGACCGC CATGCCGCG
1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
1851 TGTGCCCCGAA GTCAGGGTAA AATCGGAAGC GGGGCGAGAC GGACGGATTG
1901 TCCTGACAGT TTGCGACAAC GGCAAGGGGT TCGGCAGGGA AATGCTGCAC
1951 AATGCCTTCG AGCCGTATGT AACGGACAAA CCGGCTGGAA CGGGATTGNG
2001 ACTGCCCGTG GTGAAAAAAA TCATTGAAGA ACACGGCGGC CNCATCAGCC
2051 TGAGCAATCA GGATGCGGGC GGCGCGTNTG TCAGAAATCAT CTTGCCAAAA
2101 ACGGTAGAAA CTTATGCGTA G
  
```

This encodes a protein having amino acid sequence <SEQ ID 254>:

30  
 35  
 40

```

1 MRREFLPAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
51 LARYVILLK DRRDGVFGSQ IAKRLSGMFT LVAVLPGVFL FGVSAQFING
101 TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNAIPVQ IDXIGAASLP
151 XDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI
201 QQAGSVRDxE SIGGVLYAXG WLSAXTHNGR DYALFFRQPV PKGVAEDAVL
251 IEKARAXXXX LSYSKKGLQT FFLATLLIAS LLSIFLALVM ALYFARREVE
301 PVLSLAEGAK AVAQGDFSQT RPYLRNDEFG RLTKLFNHMT EQLSIAKEAD
351 ERNRRREEAA RHYLECVLEG LTTGVVVFDE QGCLKTFNKA AEQILGMPLT
401 PLWGSSRHGW HGVSAQSSL AEFVFAIGAA AGTDKPVHVK YAAPDDAKIL
451 LGKATVLPED NXNGVVMVID DITVLIHAQK EAAWGEVAKR LAHEIRNPLT
501 PIQLSAERLA WKLGGKLDX DAQILTRSTD TIIKQVAALK EMVEAFRNYX
551 RSPSXQLENQ DLNALIGDVL ALYEAGPCRF AAELAGEPLM MAADTTAMRQ
601 VLHNIFKNAA EAAEEADVPE VRVKSEAGQD GRIVLTVCDN GKGFGREMLH
651 NAFEPYVDK PAGTGLXLPV VKKIIEEHGG XISLSNQDAG GAXVRIILPK
701 TVETYA*
  
```

ORF64a and ORF64-1 show 96.6% identity in 706 aa overlap:

45  
 50  
 55  
 60  
 65

```

              10      20      30      40      50      60
orf64a.pep  MRREFLPAAICAVVLLYGLTAATGSTSSLADYFWWIVAFSAML VLSAVLARYVILLK
              |||
orf64-1     MRREFLPAAICAVVLLYGLTAATGSTSSLADYFWWIVAFSAML VLSAVLARYVILLK
              10      20      30      40      50      60

              70      80      90      100     110     120
orf64a.pep  DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
              |||
orf64-1     DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
              70      80      90      100     110     120

              130     140     150     160     170     180
orf64a.pep  SKSALNLAADNALGNAIPVQIDXIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIEK
              |||
orf64-1     SKSALNLAADNALGNAIPVQIDLIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIEK
              130     140     150     160     170     180

              190     200     210     220     230     240
orf64a.pep  SINPHKLDQPFPGKARWEKIQQAGSVRDLESIGGVLYAXGWLSAXTHNGR DYALFFRQPV
              |||
orf64-1     SINPHKLDQPFPGKARWEKIQQAGSVRDLESIGGVLYAXGWLSAXTHNGR DYALFFRQPV
              190     200     210     220     230     240

              250     260     270     280     290     300
  
```

-184-

5	orf64a.pep	PKGVAEDAVLIEKARAXXXLSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFVE
	orf64-1	PKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFVE
10	orf64a.pep	PVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRLTCLFNHMTQLSIAKEADERNNRREEAA
	orf64-1	PVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRLTCLFNHMTQLSIAKEADERNNRREEAA
15	orf64a.pep	RHYLECVLEGLTTGVVVVFDEQGCLKTFNKAAEQILGMPLTPLWGSSRHGWGHVSAQQSLL
	orf64-1	RHYLECVLEGLTTGVVVVFDEQGCLKTFNKAAEQILGMPLTPLWGSSRHGWGHVSAQQSLL
20	orf64a.pep	AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNNGVVMVIDDITVLIHAQK
	orf64-1	AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNNGVVMVIDDITVLIHAQK
25	orf64a.pep	EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGKLDQDAQILTRSTDTIIVKQVAALK
	orf64-1	EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGKLDQDAQILTRSTDTIIVKQVAALK
30	orf64a.pep	EMVEAFRNYXRSPSXQLENQDLNALIGDVLALYEAGPCRFAAELAGEPLTVAADTTAMRQ
	orf64-1	EMVEAFRNYXRSPSLKLENQDLNALIGDVLALYEAGPCRFAAELAGEPLTVAADTTAMRQ
35	orf64a.pep	VLHNI FKNAAEAAEADVPEVRVKSEAGQDGRIVLTVCDNGKGF GREMLHNAFEPYVTDK
	orf64-1	VLHNI FKNAAEAAEADVPEVRVKSETGQDGRIVLTVCDNGKGF GREMLHNAFEPYVTDK
40	orf64a.pep	PAGTGLXLPVVKKIEEHGGXISLSNQDAGGAXVRIILPKTVETYAX
	orf64-1	PAGTGLGLPVVKKIEEHGGGRISLSNQDAGGACVRIILPKTVKTYAX

Homology with a predicted ORF from *N.gonorrhoeae*ORF64 shows 86.6% identity over a 387aa overlap with a predicted ORF (ORF64.ng) from *N.*50 *gonorrhoeae*:

55	orf64.pep	MRRFLPIAICAXXLXXGLTAATGSTSSLADYFWWIVAFSAMPLLLVLSAVLARYVILLK	60
	orf64ng	MRRFLPIAICAVVLLYGLTAATGSTSSLADYFWWIVSFSAMPLLLVLSAVLARYVILLK	60
60	orf64.pep	DRRDGVFGSXXAKXPXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN	120
	orf64ng	DRRNGVFGSQIAKR-LSGMFTLVAVLPLGLFLFGISAQFINGTINSWFGNDTHEALERSLN	119
65	orf64.pep	LSKSALNLAADNALGNAPVQIDLIGAASLPDGMGRVLEHYAGSGFAQLALYNXASGKIE	180
	orf64ng	LSKSALDLAADNAVSNAPVQIDLIGTASLSGNMGSVLEHYAGSGFAQLALYNAASGKIE	179
65	orf64.pep	KSINPHKLDQPFPGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP	240
	orf64ng	KSINPHQFDQPLPDKEHWEKIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYALFFRQP	239

-185-

	orf64.pep	VPKGVAEADAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV	300
	orf64ng	IPENVAQDAVLIEKARAKYAELSYSKKGLQTFVLVTLIASLLSIFLALVMALYFARRFV	299
5	orf64.pep	EPVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRLTXLFNHMTQLSIAKDADERNRRREEA	360
	orf64ng	EPILSLAEGAKAVAQGDFSQTRPVLNRNDEFGRLTXLFNHMTQLSIAKEADERNRRREEA	359
10	orf64.pep	ARHYLECVLEGLTTGVVVFDEQGCCLKTFNKAAGT	394
	orf64ng	ARHYLECVLDGLTTGVVVSYP LSCCRTAVFSTCHSSPLSYF	400

An ORF64ng nucleotide sequence <SEQ ID 255> was predicted to encode a protein having amino acid sequence <SEQ ID 256>:

	1	MRRFLPIAAI	CAVLLYGLT	AATGSTSSLA	DYFWWIVSFS	AMLLLVLSAV
15	51	LARYVILLK	DRRNGVFGSQ	IAKRLSGMFT	LVAVLPGLFL	FGISAQFING
	101	TINSWFGNDT	HEALERSLNL	SKSALDLAAD	NAVSNAPVQ	IDLIGTASLS
	151	GNMGSVLEHY	AGSGFAQLAL	YNAASGKIEK	SINPHQFDQP	LPDKEHWEQI
	201	QQTGSVRSLE	SIGGVLYAQG	WLSAGTHNGR	DYALFFRQPI	PENVAQDAVL
	251	IEKARAKYAE	LSYSKKGLQT	FFLVTLIAS	LLSIFLALVM	ALYFARRFVE
20	301	PILSLAEGAK	AVAQGDFSQT	RPVLRNDEFG	RLTKLFNHMT	EQLSIAKEAD
	351	ERNRRREEAA	RHYLECVLDG	LTTGVVVSYP	LSCCRTAVFS	TCHSSPLSYF*

Further work revealed the complete gonococcal DNA sequence <SEQ ID 257>:

	1	ATGCGCCGCT	TCCTACCGAT	CGCAGCCATA	TGCGCCGTCG	TCCTGCTGTA
25	51	CGGATTGACG	GCGGCGACCG	GCAGCACCAG	TTCGCTGGCG	GATTATTTCT
	101	GGTGGATAGT	CTCGTTCAGC	GCAATGCTGC	TGCTGGTGTT	GTCCGCCGTT
	151	TTGGCACGTT	ATGTCAATATT	GCTGTTGAAA	GACAGGCGCA	ACGGCGTGTT
	201	CGGTTCCGAG	ATTGCCAAAC	GCCTTCCCG	GATGTTACAG	CTGGTCGCCG
	251	TACTGCCCGG	CTTGTTCTCG	TTCGGCATT	CCGCGCAGTT	TATCAACGGC
	301	ACGATTAAAT	CGTGGTTCGG	CAACGACACC	CACGAAGCCC	TCGAACGCAG
30	351	CCTTAATTG	AGCAAGTCCG	CACTGGATT	GGCGGCAGAC	AATGCCGTCA
	401	GCAACGCCGT	TCCCGTACAG	ATAGACCTCA	TCGGCACCGC	CTCCCTGTCTG
	451	GCTCAATATG	GCAGTGTGCT	GGAACACTAC	GCCGGCAGCG	GTTTGTCCCA
	501	GGTTGCCCTG	TACAATGCCG	CAAGCGGGAA	AATCGAAAAA	AGCATCAATC
	551	CGCACCAATT	CGACCAGCCG	CTTCCCGACA	AAGAACATTG	GGAACAGATT
35	601	CAGCAGACCG	GTTCCGTTTCG	GAGTTTGGA	AGCATAGGCG	GCGTATTGTA
	651	CGCGCAGGGA	TGGTTGTCTG	CAGGTACGCA	CAACGGGCGC	GATTACGCGC
	701	TGTTCTTCCG	CCAGCCGATT	CCCGAAAATG	TGGCACAGGA	TGCCGTTCTG
	751	ATTGAAAAGG	CGCGGGCGAA	ATATGCCGAA	TTGAGTTACA	GCAAAAAAGG
	801	TTTGCAGACC	TTTTTTCTGG	TAACCTGTCT	GATTGCCTCG	CTGCTGTCTGA
40	851	TTTTTCTTGC	GCTGGTAAATG	GCACTGTATT	TTGCCGCGCG	TTTCGTCTGAA
	901	CCCATTCTGT	CGCTTGCCGA	GGGCGCAAAG	GCGGTGGCGC	AGGGTGATTT
	951	CAGCCAGACG	CGCCCCGTAT	TGCGCAACGA	CGAGTTCGGA	CGTTTGACCA
	1001	AGCTGTTCAA	CCATATGACC	GAGCAGCTTT	CCATCGCCAA	AGAAGCAGAC
	1051	GAACGCAACC	GCCGGCGCGA	GGAAGCCGCC	CGTCACTACC	TCGAGTGCCT
45	1101	GTTGGATGGG	TTGACTACCG	GTGTGGTGGT	GTTTGACGAA	AAAGGCCGTT
	1151	TGAAAACCTT	CAACAAGGCG	GCGGAACAGA	TTTTGGGGAT	GCCGCTCGCC
	1201	CCCCGTGGG	GCAGCAGCCG	GCACGGTTG	CACGGCGTTT	CGGCGCAGCA
	1251	GTCCCTGCTT	GCCGAAGTGT	TtgccgccAT	CGGTGCGGCG	GCAGGTACCG
	1301	ACAAACCGGT	CCAGGTGGAA	TATGCCGCGC	CGGACGATGC	CAAAATCCTG
50	1351	CTGGGCAAGG	CGACGGTATT	GCCCCAAGAC	AACGGCAACG	GCGTGGTGAT
	1401	GGTGATTGAC	GACATCACCG	TGCTGATACG	CGCGCAAAAA	GAAGCCGCGT
	1451	GGGGTGAAGT	GGCGAAGCGG	CTGGCACACG	AAATCCGCAA	TCCGCTCAGC
	1501	CCCATCCAGC	TTTCCGCGCA	ACGGCTGGCG	TGGAATTTGG	GCGGGAAGCT
	1551	GGACGATCAG	GACGCGCAAA	TCCTGACGCG	TtcgACCGAC	ACCATCATCA
55	1601	AACAGgtggc	gGCGTTAAAA	GAAATGGTCG	AGGCATTCCG	CAATTACGCG
	1651	CGCGCCCCCT	CGCTCAAAC	GGAAATCAG	GATTTGAAACG	CCTTAATCGG
	1701	CGATGTTTTG	GCCCTGTACG	AAGCCGGCCC	GTGCCGTTT	GAGCGGGAAC
	1751	TTGCCGCGCA	ACCGCTGATG	ATGGCGGCGG	ATACGACCGC	CATGCGGCAG
	1801	GTGCTGCACA	ATATTTTCAA	AAATGCCGCC	GAAGCGGCGG	AAGAAGCCGA
60	1851	TATGCCCGAA	GTCAGGGTAA	AATCGGAAAC	GGGGCAGGAC	GGACGGATTG
	1901	TCCTGACGGT	TTGCGACAAC	GGCAAGGGAT	TCGGCAAGGA	AATGCTGCAC
	1951	AATGCTTTTC	AGCCGTATGT	GACGGATAAG	CCGGCGGGAA	CGGGACTGGG
	2001	TCTGCCTGTA	GTGAAAAAAA	TCATTGGAGA	ACACGGCGGC	CGCATCAGCC
	2051	TGAGCAATCA	GGATGCGGGT	GGGGCGTGTG	TCAGAATCAT	CTTGCCAAAA
65	2101	ACGGTAGAAA	CTTATGCGTA	G		

This corresponds to the amino acid sequence <SEQ ID 258; ORF64ng-1>:

```

1  MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVSFS AMLLLVLSAV
51  LARYVILLK  DRRNGVFGSQ  IAKRLSGMFT  LVAVLPGLFL  FGISAQFING
101 TINSWFGNDT HEALERSLNL  SKSALDLAAD NAVSNAVVPQ  IDLIGTASLS
5   GNMGSVLEHY AGSGFAQLAL  YNAASGKIEK SINPHQFDQP  LPDKEHWEQI
201 QQTGSVRSLE SIGGVLYAQG  WLSAGTHNGR DYALFFRQPI  PENVAQDAVL
251 IEKARAKYAE LSYSKKGLQT  FFLVTLIAS  LLSIFLALVM  ALYFARRFVE
301 PILSLAEGAK AVAQGDFSQT  RPVLRNDEFG RLTKLFNHMT  EQLSIAKEAD
351 ERNRRREEAA RHYLECVLDG  LTTGVVVFDE KGRLKTFNKA  AEQILGMPLA
10  401 PLWGSSRHGW HGVSAQQSLL  AEVFAAIGAA AGTDKPQVE  YAAPDDAKIL
451 LGKATVLPED NGNGVVMVID  DITVLIRAQK  EAAWGEVAKR LAHEIRNPLT
501 PIQLSAERLA WKLGGKLDQ  DAQILTRSTD  TIIKQVAALK  EMVEAFRNYA
551 RAPSLKENQ  DLNALIGDVL  ALYEAGPCRF  EAELAGEPLM  MAADTTAMRQ
601 VLHNIFKNAA  EAAEEADMPE  VRVKSETGQD  GRIVLTVCDN  GKGFGEMLH
15  651 NAFEPYVTDK  PAGTGLGLPV  VKKIIGEHHG  RISLSNQDAG  GACVRILPK
701 TVETYA*

```

ORF64ng-1 and ORF64-1 show 93.8% identity in 706 aa overlap:

```

20  orf64ng-1.pep  10      20      30      40      50      60
    MRRFLPIAAICAVVLLYGLTAATGSTSSLA DYFWWIVSFSAMLLLVLSAVLARYVILLK
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  MRRFLPIAAICAVVLLYGLTAATGSTSSLA DYFWWIVAFSAMLVLSAVLARYVILLK
    10      20      30      40      50      60

25  orf64ng-1.pep  70      80      90      100     110     120
    DRRNGVFGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNL
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  DRRDGVFGSQIAKRLSGMFTLVAVLPGLVFGVSAQFINGTINSWFGNDTHEALERSLNL
    70      80      90      100     110     120

30  orf64ng-1.pep  130     140     150     160     170     180
    SKSALDLAADNAVSNVAVPQIDLIGTASLSGNMGSVLEHYAGSGFAQLALYNAASGKIEK
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  SKSALNLAADNALGNVAVPQIDLIGAASLPDGMGRVLEHYAGSGFAQLALYNAASGKIEK
    130     140     150     160     170     180

35  orf64ng-1.pep  190     200     210     220     230     240
    SINPHQFDQPLPDKEHWEQIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYALFFRQPI
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  SINPHKLDQFPFGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHNGRDYALFFRQPV
    190     200     210     220     230     240

40  orf64ng-1.pep  250     260     270     280     290     300
    PENVAQDAVLIEKARAKYAE LSYSKKGLQTFFLVTLIASLLSIFLALVMALYFARRFVE
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  PKGVAEDAVLIEKARAKYAE LSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFVE
    250     260     270     280     290     300

45  orf64ng-1.pep  310     320     330     340     350     360
    PILSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTKLFNHMT EQLSIAKEADERNRRREEAA
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  PVLSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTKLFNHMT EQLSIAKEADERNRRREEAA
    310     320     330     340     350     360

50  orf64ng-1.pep  370     380     390     400     410     420
    RHYLECVLDGLTTGVVVFDEKGRLKTFNKA AEQILGMPLAPLWGSSRHGW HGVSAQQSLL
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  RHYLECVLEGLTTGVVVFDEQGCKLTFNKA AEQILGMPLT PLWGSSRHGW HGVSAQQSLL
    370     380     390     400     410     420

55  orf64ng-1.pep  430     440     450     460     470     480
    AEVFAAIGAAAGTDKPQVEYAAPDDAKILLGKATVLPEDNGNGVVMVIDDITVLIRAQK
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNGNGVVMVIDDITVLIHAQK
    430     440     450     460     470     480

60  orf64ng-1.pep  490     500     510     520     530     540
    EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDQDAQILTRSTD TIIKQVAALK
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDQDAQILTRSTD TIIKQVAALK
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```



-187-

	orf64-1	EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGKGLDEQDAQILTRSTDTIVKQVAALK	490	500	510	520	530	540
5	orf64ng-1.pep	EMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAEELAGEPLMMAADTTAMRQ	550	560	570	580	590	600
	orf64-1	EMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAEELAGEPLTVAADTTAMRQ	550	560	570	580	590	600
10	orf64ng-1.pep	VLHNIFKNAAEAAEEADMPEVRVKSETGQDGRIVLTVCDNGKGFGEMLHNAFEPYVTDK	610	620	630	640	650	660
	orf64-1	VLHNIFKNAAEAAEEADVPEVRVKSETGQDGRIVLTVCDNGKGFGEMLHNAFEPYVTDK	610	620	630	640	650	660
15	orf64ng-1.pep	PAGTGLGLPVVKKIIEHGGRISSNQDAGGACVRIILPKTVETYAX	670	680	690	700		
	orf64-1	PAGTGLGLPVVKKIIEHGGRISSNQDAGGACVRIILPKTVKTYAX	670	680	690	700		

Furthermore, ORF64ng-1 shows significant homology to a protein from *A. caulinodans*:

25	sp Q04850 NTRY AZOCA NITROGEN REGULATION PROTEIN NTRY >gi 77479 pir  S18624 ntry protein - Azorhizobium caulinodans >gi 38737 (X63841) NtrY gene product [Azorhizobium caulinodans] Length = 771 Score = 218 bits (550), Expect = 7e-56 Identities = 195/720 (27%), Positives = 320/720 (44%), Gaps = 58/720 (8%)	
	Query: 7	IAAICAVVLLYGLTAATGSTSSLADYFWWIXXXXXXXXXXXXXXXXXXRYVILLKDRRNGV 66
30	Sbjct: 35	I+A+ ++L GLT + + + R + + K R G
		ISALATFLILMGLTPVVPVTHQVVIS----VLLVNAAVLILSAMVGREIWRIRAKARAGR 90
35	Query: 67	FGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNLKSALD 126
	Sbjct: 91	AAARLHIRIVGLFAVVSVPAILVAVVASLTLDRLDRWFMSRTQEI VASSVSAQTYVR 150
40	Query: 127	LAADNAVSNAPVQIDLIGTASLSGNMGSVLEHYAG--SGFAQLALYNAASGKIEKSINP 184
	Sbjct: 151	A N + + + DL S+ Y G S F Q+ AA + ++
45	Query: 185	HQFDQPLPDKEHWEQIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYA----- 233
	Sbjct: 201	RR-DLSVVERAN-VNIGREFIVPANLAIGDATPDQPVIIYLP--NDADYVAAVPLKDYDD 256
50	Query: 234	--LFFRQPIPENVAQDAVLIEKARAKYAELSYSKGLQTFVLVXXXXXXXXXXXXXVMA 291
	Sbjct: 257	L+ + I V ++ A Y L + G+Q F + +
55	Query: 292	LYFARRFVEPILSLAEGAKAVAQGDQFSQTRPVLRLND-EFGRLTKLFNHMTQELSIXXXXX 350
	Sbjct: 317	L F++ V PI L A VA+G+ P+ R + + L + FN MT +L
60	Query: 351	XXXXXXXXXXXXHYLECVLDGLTTGVVVFDEKGRKLTFNKAAEQILGMPLAPLWGSSRHGW 410
	Sbjct: 377	+ E VL G+ GV+ D + R+ N++AE++LG L+ + RH
65	Query: 411	HGVSAQQSLLAEVFXXXXXXXXXTDKPVQVEYAAPDDAKILLGKATVLPEDNG---NGVVM 467
	Sbjct: 435	V LL E + VQ D + + V E + +G V+
70	Query: 468	VIDDITVLIRAQKEAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGKGLDDQDAQILTR 527
	Sbjct: 489	+DDIT LI AQ+ +AW +VA+R+AHEI+NPLTPIQLSAERL K G + QD +I +
75	Query: 528	STDIIKQVAALKEMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAEELAGE 587
	Sbjct: 548	TDITII+QV + MV+ F ++AR P +++QD++ +I + L G +
80	Query: 588	PLMMAA-DTTAMRQVLHNIFKNXXXXXXXXDMPEVRVK-----SETGQDGRIVLTVCD 639
	Sbjct: 608	P M A D + Q L NI KN P+VR + + G+D +V+ + D
85	Query: 608	PAMPARFDRRLVSQALTNILKNAEAEI EAVP-PDVRGQGRIRVSANRVGED--LVIDIID 664

Query: 640 NGKGFGEKMLHNAFEPYVTDKPGTGLGLPVVKKIIEHGGRISLSNQDAG-GACVRIIL 698  
 NG G +E + EPYVT + GTGLGL +V KI+ EHGG I L++ G GA +R+ L  
 Sbjct: 665 NGTGLPQESRNLLEPYVTTREKGTGLGLAIVGKIMEEHGGGIELNDAPEGRGAWIRLTL 724

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 31

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 259>:

```

1  ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
51  GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101 CTTTCCAAAT TTTCGGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC
151 TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT
201 GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCC GCCCTT TTGCTTTCCT
251 ACGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACAGG CTTGGGCGCG
301 CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCCTTAG CCAGCTTTGC
351 CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTTGTATTC AACAAATTAC
401 GCCGTCTGAA AGCGTGGTGG ATTGCAACCGA ACGCATCAAC CGTCATCGGG
451 CACGCGTTGG ATACG...
```

This corresponds to the amino acid sequence <SEQ ID 260; ORF66>:

```

1  MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFQIFGI HTTWGAFSFP
51  FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101 LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRLKAWW IAPNASTVIG
151 HALDT...
```

Further work revealed the complete nucleotide sequence <SEQ ID 261>:

```

1  ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
51  GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101 CTTTCCAAAT TTTCGGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC
151 TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT
201 GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCC GCCCTT TTGCTTTCCT
251 ACGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACAGG CTTGGGCGCG
301 CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCCTTAG CCAGCTTTGC
351 CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTTGTATTC AACAAATTAC
401 GCCGTCTGAA AGCGTGGTGG ATTGCAACCGA CCGCATCAAC CGTCATCGGC
451 AACGCCTTGG ATACGCTGGT ATTTTTCGCC GTTGCTTCT ACGCAAGCAG
501 CGATGGATT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
551 TGTTCAAAC TACCGTCTGC ACCCTCTTCT TCCTGCCCCG CTACGGCGTG
601 ATACTGAATC TGCTGACGAA AAAACTGACA ACCCTGCAAA CCAAACAGGC
651 GCAAGACCGC CCCGCGCCCT CGCTGCAAAA TCCGTAA
```

This corresponds to the amino acid sequence <SEQ ID 262; ORF66-1>:

```

1  MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFQIFGI HTTWGAFSFP
51  FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101 LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRLKAWW IAPTASTVIG
151 NALDTLVFFA VAFYASSDGF MAANWQGI AF VDYLEFKLTVC TLFFLPAYGV
201 ILNLLTKKLT TLQTKQAQDR PAPSLQNP*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o221 of *E. coli* (accession number P37619)

ORF66 and o221 protein show 67% aa identity in 155aa overlap:

```

orf66 1 MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV 60
      M F+ Q+ KALF L LFH+L+I +SNYLVQ P I G HTTWGAFSFPFIFLATDLTV
o221 1 MNVFSQTQRYKALFWLSLFHLLVITSSNYLVQLPVSILGFHTTWGAFSFPFIFLATDLTV 60

5 orf66 61 RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 120
      RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA
o221 61 RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFALAHFNLFVARIATASF MAYA 120

orf66 121 IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT 155
      +GQILD+ VFN+LR+ + WW+AP AST+ G+ DT
10 o221 121 LGQILDVHVFNRLRQSRRWLAPTASTLFGNVSDT 155

```

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF66 shows 96.1% identity over a 155aa overlap with an ORF (ORF66a) from strain A of *N.*

15 *meningitidis*:

```

                10      20      30      40      50      60
orf66.pep MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV
            |||
orf66a MYAFTAAQQQKALFWLVLFHILIIAASNYLVQFPFQISGIHTTWGAFSFPFIFLATDLTV
            |||
20                10      20      30      40      50      60

                70      80      90      100     110     120
orf66.pep RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA
            |||
25 orf66a RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA
            |||
                70      80      90      100     110     120

                130     140     150
orf66.pep IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT
            :|||
30 orf66a LGQILDIFVFNKLRLKAWWVAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGIAF
            |||
                130     140     150     160     170     180

orf66a VDYLEFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX
35                190     200     210     220

```

The complete length ORF66a nucleotide sequence <SEQ ID 263> is:

```

1 ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCTGGCTGGT
51 GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
40 101 CCTTCCAAAT TTCCGGCATC CACACCACTT GGGGCGCGTT TTCCTTTCCC
151 TTCATCTTCC TCGCCACCGA CCTGACCGTC CGCATTTTCG GTTCGCACTT
201 GGCACGGCGG ATTATCTTTT GGGTCATGTT CCCC GCCCTT TTGCTTTTCT
251 ACGTCTTTTC CGTTTGTGTC CACAACGGCA GTTGGACGGG CTTGGGCGCG
301 CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCGCTGG CAAGTTTTCG
45 351 CGCCTACGCG CTCGGACAAA TCCTTGATAT TTTTGTGTTT AACAAATTAC
401 GCCGTCTGAA AGCGTGGTGG GTTGCCCGCA CTGCATCAAC CGTCATCGGC
451 AACGCCTTAG ATACGTTGGT ATTTTTCGCC GTTGCCTTCT ACGCAAGCAG
501 CGATGGATTT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
551 TGTTCAAACT CACCGTCTGC GGTCTGTTT TCCTGCCCCG CTACGGCGTG
50 601 ATTCTGAATC TGCTGACGAA AAACTGACG ACCCTGCAAA CCAAACAGGC
551 GCAAGACCGC CCCGCGCCCT CGCTGCAAAA TCCGTAA

```

This encodes a protein having amino acid sequence <SEQ ID 264>:

```

1 MYAFTAAQQQ KALFWLVLFH ILIIAASNYL VQFPFQISGI HTTWGAFSFP
51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101 LSEFNTFVGR IALASFAAYA LGQILDIFVF NKLRLKAWW VAPTASTVIG
55 151 NALDTLVFFA VAFYASSDGF MAANWQGI AF VDYLEFKLTV GLFFLPAYGV
201 ILNLLTKKLT TLQTKQAQDR PAPSLQNP*

```

ORF66a and ORF66-1 show 97.8% identity in 228 aa overlap:

```

                10      20      30      40      50      60
orf66a.pep MYAFTAAQQQKALFWLVLFHILIIAASNYLVQFPFQISGIHTTWGAFSFPFIFLATDLTV
            |||
60 orf66-1 MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV

```

-190-

		10	20	30	40	50	60
		70	80	90	100	110	120
5	orf66a.pep	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA					
	orf66-1	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA					
		70	80	90	100	110	120
10	orf66a.pep	LGQILDIFVFNKLRLRLKAWWVAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI AF					
	orf66-1	IGQILDIFVFNKLRLRLKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI AF					
		130	140	150	160	170	180
15	orf66a.pep	VDYLFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX					
	orf66-1	VDYLFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX					
		190	200	210	220	229	
20							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF66 shows 94.2% identity over a 155aa overlap with a predicted ORF (ORF66.ng) from *N.*

*gonorrhoeae*:

25	orf66.pep	MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV	60
	orf66ng	MYALTAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV	60
	orf66.pep	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA	120
30	orf66ng	RIFGSHLARRIIFWVMFPALSLSYVFSVLFHNGSWTGLGAPSQFNTFVGRIALASFAAYA	120
	orf66.pep	IGQILDIFVFNKLRLRLKAWWIAPNASTVIGHALDT	155
	orf66ng	LGQILDIFVFDKLRLRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF	180

35 The complete length ORF66ng nucleotide sequence <SEQ ID 265> is:

```

1  ATGTACGCAT TGACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
51  GCTTTTCCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101 CCTTCCGGAT TTTCGGCATC CACACCACTT GGGCGCGGTT TTCCTTTCCC
151 TTCATCTTCC TCGCCACCGA CCTGACCGTC CGCATTTTCG GTTCGCACCTT
201 GGC GCGGCGG ATTATCTTTT GGGTGATGTT CCGCGCCCTT ttgCTTcat
251 aCGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACGGG CTGGGCGCGG
301 ctgTCCCAAT TCAACACCTT TGTCGGACGC ATCGCGCTGG CAAGTTTTCG
351 CGCCTACGCG CTCGGACAAA TCCTTGATAT TTTCGTATTC GACAAATTAC
401 GCGTCTGAA AGCGTGTTGG ATTGCCCGCG CCGCATCAAC CGTCATCGGC
451 AATGCACTGG ACACGTTAGT ATTTTTCGCC GTTGCCTTTT ACGCAAGCAG
501 CGATGAATTT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
551 TGTTCAAAC TACCGTCTGC ACCCTCTTCT TCCTGCCCGC CTACGGCGTG
601 ATACTGAATC TGCTGACGAA AAAACTGACG GCCCTGCAAA CCAAACAGGC
651 GCAAGACCGC CCGTGCCCT CGCTGCAAAA TCCGTAA

```

50 This encodes a protein having amino acid sequence <SEQ ID 266>:

```

1  MYALTAQQQ KALFRLVLFH ILIIAASNYL VQFPFRIFGI HTTWGAFSFP
51  FIFLATDLTV RIFGSHLARR IIFWVMFPAL LSYVFSVLF HNGSWTGLGA
101 PSQENTFVGR IALASFAAYA LGQILDIFVF DKLRLRLKAWW IAPAASTVIG
151 NALDTLVFFA VAFYASSDEF MAANWQGI AF VDYLFLKLTVC TLFFLPAYGV
201 ILNLLTKKLT ALQTKQAQDR PVPSLQNP*

```

An alternative annotated sequence is:

```

1  MYALTAQQQ KALFRLVLFH ILIIAASNYL VQFPFRIFGI HTTWGAFSFP
51  FIFLATDLTV RIFGSHLARR IIFWVMFPAL LSYVFSVLF HNGSWTGLGA
101 LSQENTFVGR IALASFAAYA LGQILDIFVF DKLRLRLKAWW IAPAASTVIG
151 NALDTLVFFA VAFYASSDEF MAANWQGI AF VDYLFLKLTVC TLFFLPAYGV
201 ILNLLTKKLT ALQTKQAQDR PVPSLQNP*

```

60

ORF66ng and ORF66-1 show 96.1% identity in 228 aa overlap:

```

5  orf66-1.pep  MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV 60
   orf66ng     MYALTAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV 60
10  orf66-1.pep  RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 120
   orf66ng     RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120
15  orf66-1.pep  IGQILDIFVFNKLRLKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI AF 180
   orf66ng     LGQILDIFVFDKLRRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF 180
20  orf66-1.pep  VDYLEFKLTVCTLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX 229
   orf66ng     VDYLEFKLTVCTLFFLPAYGVILNLLTKKLTALQTKQAQDRPVPSLQNPX 229

```

Furthermore, ORF66ng shows significant homology with an *E.coli* ORF:

```

20  sp|P37619|YHHQ_ECOLI HYPOTHETICAL 25.3 KD PROTEIN IN FTSY-NIKA INTERGENIC
   REGION (O221)
   >gi|1073495|pir||S47690 hypothetical protein o221 - Escherichia coli >gi|466607
   (U00039) No definition line found [Escherichia coli] >gi|1789882 (AE000423)
   hypothetical 25.3 kD protein in ftsY-nika intergenic region [Escherichia coli]
   Length = 221
25  Score = 273 bits (692), Expect = 5e-73
   Identities = 132/203 (65%), Positives = 155/203 (76%)

   Query: 1  MYALTAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV 60
   M + Q+ KALF L LFH+L+I +SNYLVQ P I G HTTWGAFSFPFIFLATDLTV
30  Sbjct: 1  MNVFSQTQRYKALFWLSLFHLLVITSSNYLVQLPVSILGFHTTWGAFSFPFIFLATDLTV 60

   Query: 61  RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120
   RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA
35  Sbjct: 61  RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFALAHFNLFVARIATASF MAYA 120

   Query: 121 LGQILDIFVFDKLRRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF 180
   LGQILD+ VF++LR+ + WW+AP AST+ GN DTL FF +AF+ S D FMA +W IA
40  Sbjct: 121 LGQILDVHVFNRRLRQSRWWLAPTASTLFGNVSDTLAFFFFIAFWRSPDAFMAEHWMEIAL 180

   Query: 181 VDYLEFKLTVCTLFFLPAYGVILN 203
   VDY FK+ + +FFLP YGV+LN
40  Sbjct: 181 VDYCFKVLISIVFFLPMYGVLLN 203

```

Based on this analysis, including the homology with the *E.coli* protein and the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 32

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 267>:

```

50  1  ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
   51  AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAyGCA GTmwrAATAT
  101  CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
  151  GTACCTAAAA ATAGTAAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
  201  TTTAACACAC AyyCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
  251  CCGCCAGCGT ATCCGCGGCC GCGGTATTGG CGGGGGTCGG CAAACTTGCC
55  301  CGCTTAGgCG CGAAATTCAG CACAAGGGCG GTtCCCTATG TCGGAACAGC
   351  CcTTTTAGCC CACGACGTAT ACGAAAcTTT CAAAGAAGAC ATACAGGCAC
   401  GAGGCTACCA ATACGACCCC GAAACCGACA AATTGTAAA AGGCTACGAA
   451  TATAGTAATT GCCTTTGGTA CGAAGACAAA AGACGTATTA ATAGAACCTA

```

501 TGGCTGCTAC GCGTTGAT..

This corresponds to the amino acid sequence <SEQ ID 268; ORF72>:

```

1  MVIKYTNLNF AKLSIIAILM MYSFEANANA VXISETVSVD TGQGAKIHKE
51 VPKNSKTYSS DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAVGKLA
5  101 RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFKVGYE
151 YSNCLWYEDK RRINRTYGCY GVD..

```

Further work revealed the complete nucleotide sequence <SEQ ID 269>:

```

1  ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
10 101 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
151 GTACCTAAAA ATAGTAA AAC TTATTCATCT GATTTAATAA AAACGGTAGA
201 TTTAACACAC ATCCCTACGG GCGCAAAGC CCGAATCAAC GCCAAAATAA
251 CCGCCAGCGT ATCCCGCGCC GGCGTATTGG CGGGGGTCGG CAAACTTGCC
301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCCTATG TCGGAACAGC
15 351 CCTTTAGCC CACGACGTAT ACGAACTTT CAAAGAAGAC ATACAGGCAC
401 GAGCTACCA ATACGACCCC GAAACCGACA AATTGCAAA GGTCTCAGGC
451 TAA

```

This corresponds to the amino acid sequence <SEQ ID 270; ORF72-1>:

```

1  MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKE
20 51 VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGKLA
101 RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFAKVS
151 *

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF72 shows 98.0% identity over a 147aa overlap with an ORF (ORF72a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf72.pep		MVIKYTNLNF AKLSIIAILM MYSFEANANA VXISETVSVD TGQGAKIHKE VPKNSKTYSS					
30 orf72a		MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKE VPKNSKTYSS					
		10	20	30	40	50	60
		70	80	90	100	110	120
35 orf72.pep		DLIKTVDLTH XPTGAKARIN AKITASVS RAGVLAVGK LARLGAKFSTR AVPYVGTALLA					
orf72a		DLIKTVDLTH IPTGAKARIN AKITASVS RAGVLAVGK LARLGAKFSTR AVPYVGTALLA					
		70	80	90	100	110	120
		130	140	150	160	170	
40 orf72.pep		HDVYETFKEDI QARGYQYDP ETDKFKVGYE YSNCLWYEDK RRINRTYGCY GVD					
orf72a		HDVYETFKEDI QARGYQYDP ETDKFAKVSGX					
		130	140	150			

The complete length ORF72a nucleotide sequence <SEQ ID 271> is:

```

45 1  ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
101 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
151 GTACCTAAAA ATAGTAA AAC TTATTCATCT GATTTAATAA AAACGGTAGA
50 201 TTTAACACAC ATCCCTACGG GCGCAAAGC CCGAATCAAC GCCAAAATAA
251 CCGCCAGCGT ATCCCGCGCC GGCGTATTGG CGGGGGTCGG CAAACTTGCC
301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCCTATG TCGGAACAGC
351 CCTTTAGCC CACGACGTAT ACGAACTTT CAAAGAAGAC ATACAGGCAC
401 GAGCTACCA ATACGACCCC GAAACCGACA AATTGCAAA GGTCTCAGGC
451 TAA

```

55 This encodes a protein having amino acid sequence <SEQ ID 272>:

-193-

```

1  MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF
51 VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAGVGKLA
101 RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFAKVS
151 *

```

5 ORF72a and ORF72-1 show 100.0% identity in 150 aa overlap:

```

10 orf72a.pep      10      20      30      40      50      60
    MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF VPKNSKTYSS
    |||||
10 orf72-1        10      20      30      40      50      60
    MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF VPKNSKTYSS

15 orf72a.pep      70      80      90     100     110     120
    DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAGVGK LARLGAKF STRAVPYV GTALLA
    |||||
15 orf72-1        70      80      90     100     110     120
    DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAGVGK LARLGAKF STRAVPYV GTALLA

20 orf72a.pep      130     140     150
    HDVYETFKED IQARGYQYDP ETDKFAKVS GX
    |||||
20 orf72-1        130     140     150
    HDVYETFKED IQARGYQYDP ETDKFAKVS GX

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

25 ORF72 shows 89% identity over a 173aa overlap with a predicted ORF (ORF72.ng) from *N. gonorrhoeae*:

```

30 orf72.pep      MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF VPKNSKTYSS 60
    || : |||||
30 orf72ng        MVTKHTNLNF AKLSIIAILM MYSFEANANA VKISETLSVD TGQGAKVHKF VPKSSNIYSS 60
    || : |||||

35 orf72.pep      DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAGVGK LARLGAKF STRAVPYV GTALLA 120
    || : |||||
35 orf72ng        DLTKAVDLTH IPTGAKARIN AKITASVSRA GVL SGVGK LVRQGA FGT RAVPYV GTALLA 120

40 orf72.pep      HDVYETFKED IQARGYQYDP ETDK FVKGYE YSNCLWYED KRRINRTY GCYGV D 173
    ||||| : |||||
40 orf72ng        HDVYETFKED IQARGCRYDP ETDK FVKGYE YANCLWYED ERRINRTY GCYGV DSSIMRL M 180

```

An ORF72ng nucleotide sequence <SEQ ID 273> was predicted to encode a protein having amino acid sequence <SEQ ID 274>:

```

40 1  MVTKHTNLNF AKLSIIAILM MYSFEANANA VKISETLSVD TGQGAKVHKF
51 VPKSSNIYSS DLTKAVDLTH IPTGAKARIN AKITASVSRA GVL SGVGKLV
101 RQGAKEGTRA VPYVGTALLA HDVYETFKED IQARGCRYDP ETDK FVKGYE
151 YANCLWYEDE RRINRTYGCY GVDSSIMRLM PDRSRFPEVK QLMESQMYRL
45 201 ARPFWNWRKE ELNKLSSLDW NNFVLNRCTF DWNGGGCAVN KGDDFRAGAS
251 FSLGRNPKYK EEMDAKKPEE ILSLKVDADP DKYIEATGYP GYSEKVEVAP
301 GTKVNMGPVT DRNGNPVQVA ATFGRDAQGN TTADVQVIPR PDLTPASAEA
351 PHAQPLPEVS PAENPANNPD PDENPGTRPN PEPDPLNPD ANPDTDGQPG
401 TSPDSPAVPD RPNGRHRKER KEGEDGGLSC DYFPEILACQ EMGKPSDRMF
50 451 HDISIPQVTD DKTWSSHNFL PSNGVCPQPK TFHVFGQRQR ASYEPLCVFA
501 EKIRFAVLLA FIIMSAFVVF GSLGGE*

```

After further analysis, the following gonococcal DNA sequence <SEQ ID 275> was identified:

```

55 1  ATGGTCACAA AACATACAAA TTGAATTTT GCGAAATTGT CGATAATTGC
51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
101 CTGAAACTCT TTCGGTTGAT ACCGGACAAG GCGCGAAAGT TCATAAGTTC
151 GTTCCTAAAT CAAGTAATAT TTATTCATCT GATTTAACAA AAGCGGTAGA
201 TTTAACGCAT ATCCCCACGG GCGCAAAGC CCGAATCAAC GCCAAAATAA
251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGT CGGGGGTCGG CAAACTTGTC
301 CGCCAAGGCG CGAAATTCGG CACAAGGGCG GTTCCCTATG TCGGAACAGC
351 CCTTTTAGCC CACGACGTAT ACGAAACTTT CAAAGAAGAC ATACAGGCAC
60 401 GAGGCTGCCG ATACGATCCC GAAACCGACA AATT

```

5 ORF72ng-1 and ORF721-1 show 89.7% identity in 145 aa overlap:

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 33

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 277>:

35 This corresponds to the amino acid sequence <SEQ ID 278; ORF73>:

Further work revealed the complete nucleotide sequence <SEQ ID 279>:

This corresponds to the amino acid sequence <SEQ ID 280; ORF73-1>:

50 1 MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAAGFA AGVLMRLRHTG  
51 LSGLLLAGAA MRSGRVSVY QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL  
101 LPEKGGAVLO AGGAENFNM NOSGRKEGFS RDDDIIEGEY TVEEPPYGGNR



151 SRNAIEHKKD E\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF73 shows 90.8% identity over a 76aa overlap with an ORF (ORF73a) from strain A of *N.*

5 *meningitidis*:

```

              10      20      30      40      50      60
orf73.pep    MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLFLMAAGFAAGVLMRLQTGLTGLLLAGAA
              |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10 orf73a     MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLFLMAATFAAGVVMRLRHTGLSGLLLAGAA
              10      20      30      40      50      60

              70
orf73.pep    MRSGBKSVYQMLWPI
              |||||:||||| ||| |
15 orf73a     MRSGRVSVYXMLWXIRYTVAAVCXMSPGFVSSVXAVLLXLPFKGGAVLQAGGAENFFNM

```

The complete length ORF73a nucleotide sequence <SEQ ID 281> is:

```

1  ATGAGATTTT TCGGTATCGG TTTTTTGGTG CTGCTGTTTT TGGAGATTAT
51  GTCGATTGTG TGGGTTGCCG ATTGGTTGGG CGGCGGTTGG ACGCTGTTTC
101 TAATGGCGGC AACCTTTGCC GCCGGCGTGG TGATGCTCAG GCATACGGGG
20  CTGTCCGGTC TTTTATTGGC GGGCGCGGCA ATGAGAAGCG GCGGGAGGGT
201 ATCCGTTTAT CANATGTTGT GGCNTATCCG TTATACGGTG GCGGCGGTGT
251 GTCNGATGAG TCCGGGATTC GTATCCTCGG TGTNGGCGGT ATTGCTGNTG
301 CTNCCGTTTA AGGGAGGTGC AGTGTTGCAG GCAGGAGGTG CGGAAAATTT
351 TPTCAACATG AACANTCGG GCAGAAAAGA NGGCNTTCC CGCGATGACG
25  401 ATATTATCGA GGGGGAATAT ACGGTTGAAG ANCCTTACGG CGGCANTCGT
451 TTCCGAAACG CCNTNGAACA CAAAAAGAC GAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 282>:

```

1  MRFFGIGFLV LLFLEIMSIW VVADWLGGGW TLFLMAATFA AGVVMRLRHTG
51  LSGLLLAGAA MRSGRVSVY XMLWXIRYTV AAVCXMSPGF VSSVXAVLLX
30  101 LPFKGGAVLQ AGGAENFFNM NXSGRKXGXS RDDDIIEGEY TVEXPYGGXR
151 FRNAXEHKKD E*

```

ORF73a and ORF73-1 show 91.3% identity in 161 aa overlap

```

              10      20      30      40      50      60
35 orf73a.pep MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLFLMAATFAAGVVMRLRHTGLSGLLLAGAA
              |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf73-1       MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLFLMAAGFAAGVLMRLRHTGLSGLLLAGAA
              10      20      30      40      50      60

              70      80      90      100     110     120
40 orf73a.pep MRSGRVSVYXMLWXIRYTVAAVCXMSPGFVSSVXAVLLXLPFKGGAVLQAGGAENFFNM
              |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf73-1       MRSGRVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLLPFKGGAVLQAGGAENFFNM
              70      80      90      100     110     120

              130     140     150     160
45 orf73a.pep NXSGRKXGXS RDDDIIEGEYTVEXPYGGXRFRNAXEHKKDEX
              |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf73-1       NQSGRKEGFSRDDDIIEGEYTV EEPYGGNRSRNAIEHKKDEX
              130     140     150     160
50

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF73 shows 92.1% identity over a 76aa overlap with a predicted ORF (ORF73.ng) from *N. gonorrhoeae*:

```

55 orf73.pep    MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLFLMAAGFAAGVLMRLQTGLTGLLLAGAA    60
              |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

-196-

```

    orf73ng      MRFFGIGFLVLLFLEIMSI VVADWLGGGWTFLMAATFAAGVLMRLRHTGLSGLLLAGAA      60
    orf73.pep    MRSGBKVSQYQMLWPI                                                    76
                  ::|||:|||||
5    orf73ng      VKSSGKVSQYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFGKGAVLQAGGAENFFNM      120

```

The complete length ORF73ng nucleotide sequence <SEQ ID 283> is:

```

      1  ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTTT TGGAAATTAT
     51  GTCGATTGTG TGGGTGCGCG ATTGGCTGGG CGGCGGTTGG AcgcTGTTC
    101  TAATGGCGGC AACCTTTGCC GCCGGTGTGC TGATGCTCAG GCATAcggGG
    151  CTGTCCGGTC TTTTATTGGC TGGCGCGGCG GTAAAAagta gtgGGAAGGT
    201  ATCTGTTTAT CagatgtTGT GGCCTATCCG TTATAcggtg gcggcggtgT
    251  GTCTGatgag tCcggGATTC GTATCCTccg tgttggCGGT ATTGCTGCTG
    301  CTGCcgttta aggGaggGgc agtggtgcag gcaggaggtg cggaaaATTT
    351  TTTCAACATg aaCcaatcgg gcagaaAaga gggatttttc cacgatgacg
    401  atattatcga gggagaatat acggttgaaa aacctgacgg cggcaatcgt
    451  tcccgaAAcg ccatcgaaca cgaaaAagac gaataA

```

This encodes a protein having amino acid sequence <SEQ ID 284>:

```

      1  MRFFGIGFLV LLFLEIMSI VVADWLGGGW TLFLMAATFA AGVLMRLRHTG
     51  LSGLLLAGAA VKSSGKVSQY QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL
    101  LPFGKGA VLQ AGGAENFFNM NQSGRKEGFF HDDDIIEGEY TVEKPDGGRN
    151  SRNAIEHEKD E*

```

ORF73ng and ORG73-1 show 93.8% identity in 161 aa overlap

```

      10      20      30      40      50      60
    orf73-1.pep MRFFGIGFLVLLFLEIMSI VVADWLGGGWTFLMAAGFAAGVLMRLRHTGLSGLLLAGAA
    orf73ng      MRFFGIGFLVLLFLEIMSI VVADWLGGGWTFLMAATFAAGVLMRLRHTGLSGLLLAGAA
      10      20      30      40      50      60
      70      80      90     100     110     120
    orf73-1.pep MRSGRVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFGKGA VLQAGGAENFFNM
    orf73ng      VKSSGKVSQYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFGKGA VLQAGGAENFFNM
      70      80      90     100     110     120
      130     140     150     160
    orf73-1.pep NQSGRKEGFSRDDDIIEGEYTV EEPYGGNRSR NAIEHKKDEX
    orf73ng      NQSGRKEGFFHDDDIIEGEYTV EKP DGGNRSR NAIEHEKDEX
      130     140     150     160

```

Based on this analysis, including the presence of a putative leader sequence and putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 34

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 285>:

```

      1  ATGTTTGT TTTCAGACGGC ATTCTT.ATG TTTCAGAAAC ATTTGCAGAA
     51  AGCCTCCGAC AGCGTCGTCG GAGGGACATT ATACGTGGTT GCCACGCCCA
    101  TCGGCAATTT GGCGGACATT ACCCTGCGCG CTTTGGCGGT ATTGCAAAAG
    151  GCG..... GCGA AGACACGCGC GTTACCGCAC AGCTTTTGAG
    201  CGCGTACGGC ATTCAGGGCA AACTCGTCAG TGTGCGCGAA CACAACGAAC
    251  GGCAGATGGC GGACAAGATT GTCGGCTATC TTTCAGACGG CATGGTTGTG
    301  GCACAGGTTT CCGATGCGGG TACGCCGGCC GTGTGCGACC CGGGCGCGAA
    351  ACTCGCCCGC CGCGTGCGTG AGGCCGGGTT TAAAGTCGTT CCCGTGCTGG
    401  GCGCAAC.GC GGTGATGGCG GCTTTGAGCG TGGCCGGTGT GGAAGGATCC
    451  GATTTTATT TCAACGGTTT TGTACCGCCG AAATCGGGAG AACGCAGGAA

```

501 ACTGTTTGCC AAATGGGTGC GGGCGGCGTT TCCTATCGTC ATGTTTGAAA  
 551 CGCCGCACCG CATCGGTGCA GCGCTTGCCG ATATGGCGGA ACTGTTCCCC  
 601 GAACGCCGAT TAATGCTGGC GCGCGAAATT ACGAAAACGT TTGAAACGTT  
 651 CTTAAGCGGC ACGGTTGGGG AAATTCAGAC GGCATTGTCT GCCGACGGCG  
 701 ACCAATCGCG CGGCGAGATG GTGTTGGTGC TTTATCCGGC GCAGGATGAA  
 751 AAACACGAAG GCTTGTCGGA GTCCGCGCAA AACATCATGA AAATCCTCAC  
 801 AGCCGAGCTG CCGACCAAAC AGGCGGCGGA GCTTGCTGCC AAAATCACGG  
 851 GCGAGGGAAA GAAAGCTTTG TACGAT..

This corresponds to the amino acid sequence <SEQ ID 286; ORF75>:

1 MFVFQTAFXM FQKHLQKASD SVVGGTLYVV ATPIGNLADI TLRALAVLQK  
 51 A...AEDTR VTAQLLSAYG IQGKLVSVRE HNERQMADKI VGYLSDGMV  
 101 AQVSDAGTPA VCDPGAKLAR RVRGAGFKV PVVGAXAVMA ALSVAGVEGS  
 151 DFYFNGFVPP KSGERRKLFA KVVRAAFPIV MFETPHRIGA ALADMAELFP  
 201 ERRMLLAREI TKTFTFSLG TVGEIQTALS ADGDQSRGEM VLVLYPAQDE  
 251 KHEGLSESAQ NIMKILTAEL PTKQAAELAA KITGEGKKAL YD..

Further work revealed the complete nucleotide sequence <SEQ ID 287>:

1 ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC  
 51 ATTTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC  
 101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG  
 151 CGCGTTACCG CACAGCTTTT GAGCGCGTAC GGCATTACAG GCAAACCTCGT  
 201 CAGGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTCGGCT  
 251 ATCTTTCAGA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG  
 301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCGG  
 351 GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA  
 401 GCGTGGCCGG TGTGAAGGA TCCGATTTTT ATTTCAACGG TTTTGTACCG  
 451 CCGAAATCGG GAGAACGCAG GAAACTGTTT GCCAAATGGG TCGGGGCGGC  
 501 GTTTCCTATC GTCATGTTTG AAACGCCGCA CCGCATCGGT GCGACGCTTG  
 551 CCGATATGGC GGAAGTGTTT CCGGAACGCC GATTAATGCT GGCGCGCGAA  
 601 ATTACGAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA  
 651 GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGGCGAG ATGGTGTGTTG  
 701 TGCTTTATCC GGCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG  
 751 CAAACATCA TGAATATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC  
 801 GGAGCTTGCT GCCAAAATCA CGGGCGAGGG AAAGAAAGCT TTGTACGATC  
 851 TGGCTCTGTC TTGAAAAAC AAATAG

35 This corresponds to the amino acid sequence <SEQ ID 288; ORF75-1>:

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT  
 51 RVTAQLLSAY GIQKLVSVR EHNERQMADK IVGYLSDGMV VAQVSDAGTP  
 101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVEG SDFYFNGFVP  
 151 PKSGERRKLF AKVVRAAFPI VMFETPHRIG ATLADMAELF PERRMLLARE  
 201 ITKTFTFELS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA  
 251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF75 shows 95.8% identity over a 283aa overlap with an ORF (ORF75a) from strain A of *N.*

45 *meningitidis*:

		10	20	30	40	50	60
orf75.pep	MFVFQTAFXM	FQKHLQKASD	SVVGGTLYVV	ATPIGNLADI	ITLRALAVLQ	KAXXXXAEDTR	
orf75a		MFQKHLQKASD	SVVGGTLYVV	ATPIGNLADI	ITLRALAVLQ	KADIICAEDTR	
50			10	20	30	40	50
		70	80	90	100	110	120
orf75.pep	VTAQLLSAYGI	QKLVSVREH	NERQMADKI	VGYLSDGMV	VAQVSDAGT	PAVCDPGAKLAR	
orf75a	VTAQLLSAYGI	QKLVSVREH	NERQMADKI	VGYLSDGMV	VAQVSDAGT	PAVCDPGAKLAR	
55		60	70	80	90	100	110
		130	140	150	160	170	180
orf75.pep	RVREAGFKV	VPVVGAXAV	MAALSVAG	VEGSDFYF	NFGFVPP	KSGERRKL	FAKVVRAAFPIV

	orf75a	RVREVGFRKVVPPVVGASAVMAALSVAGVAGSDFYFNGFVPPKSGERRKLFAKWVRVAFPVV 120       130       140       150       160       170
5	orf75.pep	MFETPHRIGAAALADMAELFPERRLMLAREITKTFFETFLSGTVGEIQTALSADGDQSRGEM 190       200       210       220       230       240
	orf75a	MFETPHRIGATLADMAELFPERRLMLAREITKTFFETFLSGTVGEIQTALAADGNQSRGEM 180       190       200       210       220       230
10	orf75.pep	VLVLYPAQDEKHGLSESAQNIMKILTAELPTKQAELAAKITGEGKKALYD 250       260       270       280       290
	orf75a	VLVLYPAQDEKHGLSESAQNIMKILTAELPTKQAELAAKITGEGKKALYDLALSWKNK 240       250       260       270       280       290
15	orf75a	X

20	1	ATGTTTCAGA	AACATTTGCA	GAAAGCCTCC	GACAGCGTCG	TCGGAGGGAC
	51	ATTATACGTG	GTTGCCACGC	CCATCGGCAA	TTTGGCGGAC	ATTACCTTGC
	101	GCGCTTTGGC	GGTATTGCAA	AAGGCGGACA	TCATCTGTGC	CGAAGACACG
	151	CGCGTTACCG	CGCAGCTTTT	GAGCGCGTAC	GGCATTTACG	GCAAACTCGT
25	201	CACGCGTGCG	GACACAAACG	AACGGCAGAT	GGCGGACAAG	ATTGTGCGCT
	251	ATCTTTCAGA	CGGCATGGTT	GTGGCACAGG	TTTCCGATGC	GGGTACGCCG
	301	GCCGTGTGCG	ACCCGGGCGC	GAAACTCGCC	CGCCGCGTGC	GTGAGGTGCG
	351	GTTTAAAGTT	GTCCCTGTTG	TCGGCGCAAG	CGCGGTGATG	GCGGCTTTGA
30	401	GTTTGGCTGG	TGTGGCGGGA	TCCGATTTTT	ATTTCAACGG	TTTTGTACCG
	451	CCGAAATCGG	GCGAACGTAG	GAAATTGTTT	GCCAAATGGG	TGCGGGTGGC
	501	GTTTCCCGTC	GTGATGTTTG	AAACGCCGCA	CCGCATCCGG	GCGACGCTTG
	551	CCGATATGGC	GGAACGTGTC	CCCGAACGCC	GATTAATGTC	GGCGCGCGAA
35	601	ATCACGAAAA	GTTTGAAC	GTTCTTAAGC	GGCACGGTTG	GGGAATTCAA
	651	GACGGCATTG	GCGGCGGACG	GCAACCAATC	GCGCGGCGAG	ATGGTGTTGG
	701	TGCTTTATCC	GGCGCAGGAT	GAAAAACACG	AAGGCTTGTC	CGAGTCCGCG
	751	CAAAACATCA	TGAAAATCCT	CACAGCCGAG	CTGCCGACCA	AACAGGCGGC
35	801	GGAGCTTGCC	GCCAAAAATCA	CGGGCGGAGG	AAAAAAAGCT	TTGTACGATC
	851	TGCGACTGTC	TTGGAAAAAC	AAATGA		

40

1	MFQKHLQKAS	DSVVGGTLYV	VATPIGNLAD	ITLRALAVLQ	KADIICAEDT
51	RVTAQLLSAY	GIQGKLVSVR	EHNERQMA DK	IVGYLSDGMV	VAQVSDAGTP
101	AVCDPGAKLA	RRVREVGFKV	<u>VPVVGASAVM</u>	AALSVA GVAG	SDFYFNGFVP
151	PKSGERRKLF	AKWVRVAFPV	VMFETPHRIG	ATLADMAELF	PERRLMLARE
201	ITKTFETFLS	GTVGEIQTAL	AADGNQSRGE	MVLVLVPAQD	EKKEGLSESA
251	QNIMKILTAE	LPTKQAAELA	AKITGEGKKA	LYDLALSWKN	K*

45			10	20	30	40	50	60
	orf75a.pep	MFQKHLQKASDSVVG	GTLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQ	LLSAY				
	orf75-1	MFQKHLQKASDSVVG	GTLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQ	LLSAY				
50			10	20	30	40	50	60
	orf75a.pep	GIQGKLVSVREHN	ERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREVG	FKV				
	orf75-1	GIQGKLVSVREHN	ERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAG	FKV				
55			70	80	90	100	110	120
	orf75a.pep	GIQGKLVSVREHN	ERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREVG	FKV				
	orf75-1	GIQGKLVSVREHN	ERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAG	FKV				
60			130	140	150	160	170	180
	orf75a.pep	VPVVGASAVMAALS	VAGVAGSDFYFNGFVPPKSGERRKLF	AKWVRVAF	FPVVMFETPHRIG			
	orf75-1	VPVVGASAVMAALS	VAGVEGSDFYFNGFVPPKSGERRKLF	AKWVRAAF	PIVMFETPHRIG			
			130	140	150	160	170	180
65	orf75a.pep m	ATLADMAELF	PERRRLMLAREITKT	FETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQD				

orf75-1		ATLADMAELFPERRRLMLAREITKTFETFLSGTVGEIQTALSADGNQSRGEMVLVLYPAQD	
		190 200 210 220 230 240	
		250 260 270 280 290	
5	orf75a.pep	EKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYDLALSWKNKX	
	orf75-1	EKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYDLALSWKNKX	
		250 260 270 280 290	
10 Homology with a predicted ORF from <i>N.gonorrhoeae</i>			
ORF75 shows 93.2% identity over a 292aa overlap with a predicted ORF (ORF75.ng) from <i>N.gonorrhoeae</i> :			
	orf75.pep	MFVFQTAFFMFQKHLQKASDSVVGGLYVVATPIGNLADITLRALAVLQKA----AEDTR	56
15	orf75ng	MSVFQTAFFMFQKHLQKASDSVVGGLYVVATPIGNLADITLRALAVLQKADIICAEDTR	60
	orf75.pep	VTAQLLSAYGIQGLVSVREHNERQMDKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLAR	116
	orf75ng	VTAQLLSAYGIQGLVSVREHNERQMDKIVGYLSDGLVVAQVSDAGTPAVCDPGAKLAR	120
20	orf75.pep	RVREAGFKVVPVVGASAVMAALSVAGVEGSDFYFNGFVPPKSGERRKLFKQVWRAAFPV	176
	orf75ng	RVREAGFKVVPVVGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFKQVWRAAFPV	180
25	orf75.pep	MFETPHRIGALADMAELFPERRRLMLAREITKTFETFLSGTVGEIQTALSADGNQSRGEM	236
	orf75ng	MFETPHRIGATLADMAELFPERRRLMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEM	240
30	orf75.pep	VLVLYPAQDEKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYD	288
	orf75ng	VLVLYPAQDEKHEGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKALYDLALSWKNK	300

An ORF75ng nucleotide sequence <SEQ ID 291> was predicted to encode a protein having amino acid sequence <SEQ ID 292>:

35	1	MSVFQTAFFM	FQKHLQKASD	SVVGGLYVV	ATPIGNLADI	TLRALAVLQK
	51	ADIICAEDTR	VTAQLLSAYG	IQGLVSVRE	HNERQMDKV	IGFLSDGLVV
	101	AQVSDAGTPA	VCDPGAKLAR	RVREAGFKVV	PVVGASAVMA	ALSVAGVAES
	151	DFYFNGFVPP	KSGERRKLFA	KWVRAAFPVV	MFETPHRIGA	TLADMAELFP
	201	ERRLMLAREI	TKTFETFLSG	TVGEIQTALA	ADGNQSRGEM	VLVLYPAQDE
40	251	KHEGLSESAQ	NAMKILAAEL	PTKQAAELAA	KITGEGKKAL	YDLALSWKNK
	301	*				

After further analysis, the following gonococcal DNA sequence <SEQ ID 293> was identified:

45	1	ATGTTTCAGA	AACACTTGCA	GAAAGCCTCC	GACAGCGTCG	TCGGAGGGAC
	51	ATTATACGTG	GTTGCCACGC	CCATCGGCAA	TTTGGCAGAC	ATTACCCTGC
	101	GCGCTTTGGC	GGTATTGCAA	AAGGCGGACA	TCATTGTGTC	CGAAGACACG
	151	CGCGTTACTG	CGCAGCTTTT	GAGCGCGTAC	GGCATTTCAGG	GCAGGTGGT
	201	CAGTGTGCGC	GAACACAACG	AGCGGCAGAT	GGCGGACAAG	GTAATCGGTT
50	251	TCCTTTCAGA	CGGCCTGGTT	GTGGCGCAGG	TTTCCGATGC	GGGTACGCCG
	301	GCGTGTGCG	ACCCGGGCGC	GAACTCGCC	CGCCGCGTGC	GCGAAGCAGG
	351	GTTCAAAGTC	GTTCCCGTCG	TGGGCGCAAG	CGCGGTAATG	GCGGCGTTGA
	401	GTGTGGCCGG	TGTGGCGGAA	TCCGATTTTT	ATTTCAACGG	TTTTGTACCG
	451	CCGAAATCGG	GCGAACGTAG	GAAATTGTTT	GCCAAATGGG	TGCGGGCGGC
55	501	ATTTCCTGTC	GTCATGTTTG	AAACGCCGCA	CCGAATCGGG	GCAACGCTTG
	551	CCGATATGGC	GGAATTGTTT	CCCGAACGCC	GTCTGATGCT	GGCGCGGAA
	601	ATCACGAAAA	CCTTTGAAAC	GTTCTTAAGC	GGCACGGTTG	GGGAAATTCA
	651	GACGGCATTG	GCGGCGGACG	GCAACCAATC	GCGCGGCGAG	ATGGTGTGG
	701	TGCTTTATCC	GCGCGAGGAT	GAAAAACACG	AAGGCTTGTC	CGAGTCTGCG
60	751	CAAAATGCGA	TGAAAATCCT	TGCGGCCGAG	CTGCCGACCA	AGCAGGCGGC
	801	GGAGCTTGCC	GCCAAGATTA	CAGGTGAGGG	CAAAAAGGCT	TTGTACGATT
	851	TGGCACTGTC	GTGGAAAAAC	AAATGA		

This corresponds to the amino acid sequence <SEQ ID 294; ORF75ng-1>:

-200-

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT  
 51 RVTAQLLSAY GIQGRVSVR EHNERQMA DK VIGFLSDGLV VAQVSDAGTP  
 101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVAE SDFYFNGFVP  
 151 PKSGERRKLF AKWVRAAFV VMFETPHRIG ATLDMAELF PERRLMLARE  
 201 ITKTFETFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA  
 251 QNAMKILAAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K\*

ORF75ng-1 and ORF75-1 show 96.2% identity in 291 aa overlap:

		10	20	30	40	50	60
10	orf75-1.pep	MFQKHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAY					
	orf75ng-1	MFQKHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAY					
		10	20	30	40	50	60
		70	80	90	100	110	120
15	orf75-1.pep	GIQGRVSVREHNERQMA DKIVGYLS DGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKV					
	orf75ng-1	GIQGRVSVREHNERQMA DKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKV					
		70	80	90	100	110	120
20		130	140	150	160	170	180
	orf75-1.pep	VPVVGASAVMAALSVAGVEGSDFYFNGFVPPKSGERRKLF AKWVRAAFVVMFETPHRIG					
	orf75ng-1	VPVVGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLF AKWVRAAFVVMFETPHRIG					
		130	140	150	160	170	180
25		190	200	210	220	230	240
	orf75-1.pep	ATLDMAELFPERRLMLAREITKTFETFLSGTVGEIQTALSADGNQSRGEMVLVLYPAQD					
	orf75ng-1	ATLDMAELFPERRLMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQD					
30		190	200	210	220	230	240
		250	260	270	280	290	
	orf75-1.pep	EKHEGLSESAQNIMKILTAE LPTKQAAELAAKITGEGKKALYDLALSWKNKX					
35	orf75ng-1	EKHEGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKALYDLALSWKNKX					
		250	260	270	280	290	

Furthermore, ORG75ng-1 shows significant homology to a hypothetical *E.coli* protein:

sp|P45528|YRAL\_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION (F286)  
 40 >gi|606086 (U18997) ORF\_f286 [Escherichia coli]  
 >gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region [Escherichia coli] Length = 286  
 Score = 218 bits (550), Expect = 3e-56  
 Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)  
 45 Query: 4 KHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQ 63  
 K Q A +S G LY+V TPIGNLADIT RAL VLQ D+I AEDTR T LL +GI  
 Sbjct: 2 KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVLQAVDLIAAEDTRHTGLLLQHFGIN 59  
 50 Query: 64 GRLVSVREHNERQMA DKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123  
 RL ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG L R REAG +VVP+  
 Sbjct: 60 ARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPGYHLVRTCREAGIRVPL 119  
 55 Query: 124 VGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLF AKWVRAAFVVMFETPHRIGATL 183  
 G A + ALS AG+ F + GF+P KS RR ++ +E+ HR+ +L  
 Sbjct: 120 PGPCAAITALSAGLPSDRFCYEGFLPAKSKGRRDALKAEAEPRTLIFYESTHRLLDL 179  
 Query: 184 ADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEK 242  
 D+ + E R ++LARE+TKT+ET VGE+ + D N+ +GEMVL++ +  
 60 Sbjct: 180 EDIVAVLGESRYVVLARELTKTWETIHGAPVGELLAWVKEDENRRKGEMVLIV-EGHKAQ 238  
 Query: 243 HEGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKALYDLAL 286  
 E L A + +L AELP K+AA LAA+I G K ALY AL  
 65 Sbjct: 239 EEDLPADALRTLALLQAEPLPKAAALAAEIHGVKKNALYKYAL 282

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 35

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 295>:

```

1  ATGAAACAGA AAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
51  TTTTGCGGCA GC.AAAGCAC CCGAAATCGA CCCGGCTTTG .....
//
651 ..... ..GAGTTGG TCAGAAACCA GTTGGAGCAG GGTTTGAGAC
10 701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGGAAGA AAACGGTGTC
751 AAACCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 296; ORF76>:

```

1  MKQKKTAAAV IAAMLAGFAA XKAPEIDPAL .....
//
15 201 ..... ELVRNQLEQG LRQEKARLKI DALLEENGVK
251 P*

```

Further work revealed the complete nucleotide sequence <SEQ ID 297>:

```

1  ATGAAACAGA AAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
51  TTTTGCGGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
20 101 TGGTGGCGCA GATCATGCGAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA
151 AAACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGCC GGCTACAAAC
201 TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
251 AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
301 GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAGACGAGCT
25 351 GCACAAAGTTT TACGAACAGC AAATCCGCAT GATCAAATTG CAGCAGGTCA
401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
451 GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC
501 TTTTGACGGT TTCATTATGG CGCAGCAGCT TCCCGAGCCG CTGGCTTCGC
551 AGTTTGCCGC GATGAATCGG GCGCAGCTTA CCCGCGATCC GGTCAAATTG
30 601 GGCGAACGCT ATTATCTGTT CAAACTCAGC GAGGTCGGGA AAAACCCCGA
651 CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGGAGCAG GGTTTGAGAC
701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGGAAGA AAACGGTGTC
751 AAACCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 298; ORF76-1>:

```

35 1  MKQKKTAAAV IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ
51  KPDGQAIRND AVRRLOTLEV LKNRALKEGL DKDKDVQNRF KIAEASFYAE
101 EYVRFLERSE TVSEDELHKF YEQQIRMIKL QQVSFATEEE ARQAQQLLLK
151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAMNR GDVTRDPVKL
40 201 GERYLFEKLS EVGKNPDAQP FELVRNQLEQ GLRQEKARLK IDALLEENG
251 KP*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF76 shows 96.7% identity over a 30aa overlap and 96.8% identity over a 31aa overlap with an ORF (ORF76a) from strain A of *N. meningitidis*:

```

45 orf76.pep      10      20      30
      MKQKKTAAAVIAAMLAGFAAXKAPEIDPAL
      |||||
orf76a      10      20      30      40      50      60
      MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND
50      10      20      30      40      50      60
           //
           70      80      90

```

```

orf76.pep                                XELVRNQLEQGLRQEKARLKDALLEENGVKPX
|||||||||||||||||||||:|||||||
orf76a      DVTRDPVKLGERYYLFKLSEVGKNPDAQPFELVRNQLEQGLRQEKARLKDALLEENGVKPX
              200      210      220      230      240      250

```

5 The complete length ORF76a nucleotide sequence <SEQ ID 299> is:

```

1  ATGAAACAGA AAAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTGGCAGG
51  TTTTGCGGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
101 TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA
151 AAACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGTC GGCTGCAAAC
201 TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
251 AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
301 GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAAGCGCACT
351 GCGTCAGTTT TATGAGCGGC AAATCCGCAT GATCAAATTG CAGCAGGTCA
401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
451 GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC
501 TTTTGACGGT TTCATTATGG CGCAGCAGCT TCCCAGACCG CTGGCTTCGC
551 AGTTTGACAGC GATGAATCGG GGCGACGTTA CCCGCGATCC GGTCAAATTG
601 GGCGAACGCT ATTATCTGTT CAAACTCAGC GAGGTCGGGA AAAACCCCGA
651 CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGAACAA GGTGTGAGAC
701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCA TTTTGAAGA AAACGGTGTC
751 AAACCGTAA

```

This encodes a protein having amino acid sequence <SEQ ID 300>:

```

1  MKQKKTA AAV IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAESQSQ
51  KPDGQAIRND AVRRRLQTLV LKNRALK EGL DKDKDVQNR F KIAEASFYAE
101 EYVRFLERSE TVSESALRQF YERQIRMI KL QQVSFATEEE ARQAQQLLLK
151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAAMNR GDVTRDPVKL
201 GERYYLFKLS EVGKNPDAQP FELVRNQLEQ GLRQEKARLK IDAILEENG V
251 KP*

```

ORF76a and ORF76-1 show 97.6% identity in 252 aa overlap:

```

30  orf76a.pep      10      20      30      40      50      60
      MKQKKTA AAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAESQSQKPDGQAIRND
      orf76-1      10      20      30      40      50      60
      MKQKKTA AAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAESQSQKPDGQAIRND

35  orf76a.pep      70      80      90      100     110     120
      AVRRRLQTLV LKNRALK EGLDKDKDVQNR FKIAEASFYAE EYVRFLERSETVSESALRQF
      orf76-1      70      80      90      100     110     120
      AVRRRLQTLV LKNRALK EGLDKDKDVQNR FKIAEASFYAE EYVRFLERSETVSEDELHKF

40  orf76a.pep      130     140     150     160     170     180
      YERQIRMIKLQ QVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
      orf76-1      130     140     150     160     170     180
      YEQQIRMIKLQ QVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP

45  orf76a.pep      190     200     210     220     230     240
      LASQFAAMNR GDVTRDPVKLG ERYYLFKLSEVGKNPDAQPFELVRNQLEQGLRQEKARLK
      orf76-1      190     200     210     220     230     240
      LASQFAAMNR GDVTRDPVKLG ERYYLFKLSEVGKNPDAQPFELVRNQLEQGLRQEKARLK

50  orf76a.pep      250
      IDAILEENG VKPX
      orf76-1      250
      IDALLEENG VKPX

```

# 60 Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF76 and a predicted ORF (ORF76.ng) from *N. gonorrhoeae* of the N- and C-termini show 96.7 % and 100% identity in 30 and 31 overlap, respectively:



-203-

```

    orf76.pep    MKQKKTAAAVIAAMLGFAAXKAPEIDPAL                      30
                |||||||
    orf76ng      MKQKKTAAAVIAAMLGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQRPDQQAIRND  60
                //
5    orf76.pep                                ELVRNQLEQGLRQEKAARKIDALLEENGVKP  251
                |||||||
    orf76ng      VTRNPVKLGERYLLFKLGAVGKNPDAQPFELVRNQLEQGLRQEKAARKIDALLEENGVKP  251

```

The complete length ORF76ng nucleotide sequence <SEQ ID 301> is:

```

10      1  ATGAAACAGA AAAAGACCGC TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
      51  TTTTTCGGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
    101  TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA
    151  AGACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGCC GGCTGCAAAC
    201  TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
    251  AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
    15  301  GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAAGCGCACT
    351  GCGTCAGTTT TATGAGCGGC AAATCCGCAT GATCAAATTG CAGCAGGTCA
    401  GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
    451  GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC
    501  GTTCGACGGT TTCATTATGG CGCAGCAGCT TCCCGAGCCG CTGGCTTcgc
    20  551  agtttgCCGG TATGAACCGT GGCGACGTTA CCCGCAATCC GGTCAAATTG
    601  GGCGAACGCT ATTACCTGTT CAAACTCGGC GCGGTCGGGA AAAACCCCGA
    651  CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGGAACAA GGTTTGAGGC
    701  AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGGAaga Aaacggtgtc
    751  AaacCGTAA

```

25 This encodes a protein having amino acid sequence <SEQ ID 302>:

```

30      1  MKQKKTAAAV IAAMLGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ
    51  RPDQQAIRND AVRRLQTLV LKNRALKEGL DKDKDVQNRK KIAEASFYAE
    101  EYVRFLERSE TVSESALRQF YERQIRMIKL QQVSFATEEE ARQAQQLLLK
    151  GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAGMNR GDVTRNPVKL
    201  GERYLLFKLG AVGKNPDAQP FELVRNQLEQ GLRQEKAARK IDALLEENG
    251  KP*

```

ORF76ng and ORF76-1 show 96.0% identity in 252 aa overlap

```

35      10      20      30      40      50      60
    orf76-1.pep  MKQKKTAAAVIAAMLGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDQQAIRND
                |||||||
    orf76ng      MKQKKTAAAVIAAMLGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQRPDQQAIRND
                10      20      30      40      50      60

    40      70      80      90      100     110     120
    orf76-1.pep  AVRRLQTLV LKNRALKEGLDKDKDVQNRFKIAEASFYAEYVRFLERSETVSEDELHKF
                |||||||
    orf76ng      AVRRLQTLV LKNRALKEGLDKDKDVQNRFKIAEASFYAEYVRFLERSETVSESALRQF
                70      80      90      100     110     120

    45      130     140     150     160     170     180
    orf76-1.pep  YEQQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
                ||:|||||
    orf76ng      YERQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
                130     140     150     160     170     180

    50      190     200     210     220     230     240
    orf76-1.pep  LASQFAAMNRGDVTRDPVKLGERYLLFKLSEVGKNPDAQPFELVRNQLEQGLRQEKAARK
                |||||:|||||:|||||
    orf76ng      LASQFAGMNRGDVTRNPVKLGERYLLFKLGAVGKNPDAQPFELVRNQLEQGLRQEKAARK
                190     200     210     220     230     240

    55      250
    orf76-1.pep  IDALLEENGVKPX
                |||||||
    60    orf76ng      IDALLEENGVKPX
                250

```

Furthermore, ORF76ng shows significant homology to a *B.subtilis* export protein precursor:

-204-

```

sp|P24327|PRSA_BACSU PROTEIN EXPORT PROTEIN PRSA PRECURSOR >gi|98227|pir||S15269
33K lipoprotein - Bacillus subtilis >gi|39782 (X57271) 33kDa lipoprotein
[Bacillus subtilis]
>gi|2226124|gnl|PID|e325181 (Y14077) 33kDa lipoprotein [Bacillus subtilis]
5 >gi|2633331|gnl|PID|e1182997 (Z99109) molecular chaperonin [Bacillus subtilis]
Length = 292
Score = 50.4 bits (118), Expect = 1e-05
Identities = 48/199 (24%), Positives = 82/199 (41%), Gaps = 32/199 (16%)

10 Query: 70 VLKNRALKEGLDK-----DKDVQNRFKIAEASF-----YAEYVRFLESETVSE 114
VL ++ LDK DK++ N+ K + Y ++Y++ + E +++
Sbjct: 53 VLTQLVQEKVLDKKYKVSDEIDNKLKEYKTQLGDQYTALEKQYGKDYLKEQVKYELLTQ 112

Query: 115 SA-----LRQFYERQIRMIKLQQVVSFATEEEARQAQQLLLKGLSFEGLMKRYPN 163
A +++++E I+ + A ++ A + ++ L KG FE L K Y
15 Sbjct: 113 KAAKDNIKVTADADIKEYWEGLKGKIRASHILVADKKTAEVEKKLKKGEKFEDLAKEYST 172

Query: 164 DEQAFDG-----FIMAQQLPEPLASQFAAMNRGDVTRDPVKLGERYYLFLKSEVGKNPDA 218
D A G F Q+ E + + G+V+ DPVK Y++ K +E D
20 Sbjct: 173 DSSASKGGDLGWFAKEGQMDETFSKAAFKLTGEVS-DPVKTQYGYHIKKTEERGKYDD 231

Query: 219 QPFELVRNQLEQGLRQEKA 237
EL LEQ L A
25 Sbjct: 232 MKKELKSEVLEQKLNDA 250

```

Based on this analysis, including the presence of a putative leader sequence and a RGD motif in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF76-1 (27.8kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 10A shows the results of affinity purification of the His-fusion protein, Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 10B), ELISA (positive result), and FACS analysis (Figure 10C). These experiments confirm that ORF76-1 is a surface-exposed protein, and that it is a useful immunogen.

### Example 36

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 303>:

```

1 ATGAAAAAAT CTTTCCTTAC GCTTGTCTG TATTCGTCTT TACTTACCGC
51 CAGCGAAATT GCCTTACCCC TTGGAATTGG GGATTGAAAC CTTACCGGCG
101 GCAAAAATTG CGGAAACGTT TGCGCTGACA TTTGTGATTG CTGCGCTGTA
40 151 TCTGTTTGCG CGTAATAAGG TGACGCGTTT GTTGATTGCG GTGTTTTTTG
201 CGTTCAGCAT TATTGCCAAC AATGTGCATT ACGCGGATTA TCAAAGCTGG
251 ATGACG.... //
1201 ..... CAAACCGTAT TCGAGCAGCT GCAAAAGACT CCTGACGGCA
45 1251 ACTGGCTGTT TGCCTATACC TCCGATCATG GCCAGTATGT TCGCCAAGAT
1301 ATCTACAATC AAGGCACGGT GCAGCCCGAC AGCTATCTCG TGCCGCTAGT
1351 GTTGTACAGC CCGGATAAGG CCGTGCAACA GGCTGCCAAC CAGGCTTTTG
1401 CGCCTTGCGA GATTGCCTTC CATCAGCAGC TTTCAACGTT CCTGATTCAC
1451 ACGTTGGGCT ACGATATGCC GGTTCAGGT TGTGCGGAAG GCTCGGTAAC
50 1501 GGGCAACCTG ATTACGGGTG ATGCAGGCAG CTTGAACATT CGCGACGGCA
1551 AGCGGAATA TGTTTATCCG CAATGA

```

This corresponds to the amino acid sequence <SEQ ID 304; ORF81>:

```

1 MKKSFLTLVL YSLLTASEI AYPLELGIET LPAAKIAETF ALTFVIAALY

```

Further work revealed the complete nucleotide sequence <SEQ ID 305>:

This corresponds to the amino acid sequence <SEQ ID 306; ORF81-1>:

Computer analysis of this amino acid sequence gave the following results:

**Homology with a predicted ORF from *N.meningitidis* (strain A)**

ORF81 shows 84.7% identity over a 85aa overlap and 99.2% identity over a 121aa overlap with an ORF (ORF81a) from strain A of *N. meningitidis*:

```

55          10      20      30      40      50      60
   orf81.pep  MKKSFLTLVLYSSLLTASEIAYPLELGIETLPAAKIAETFALTFVIAALYLFARNKVTRL
              |||||::| ||||| ||||| : : ||||| : ||||| ||||| ||||| ||||| : ||
   orf81a     MKKSFLVFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFVIAALYLFARYKATRL
              10      20      30      40      50      60
60          70      80

```

-206-

```

5  orf81.pep      LIAVFFAFSIIANNVHYADYQSWMT
    ||||||||||||||||| ||||:|
    orf81a       LIAVFFAFSIIANNVHYAVYQSWITGINYWLMLKEITEVGGAGASMLDKLWLPALWGVLE
                  70      80      90      100      110      120
                        //

    orf81.pep                      120      130      140
                                QTVFEQLQKTPDGNWLFAYTSDHGQYVRQD
    orf81a       IPHANGLEQISGGDIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLFAYTSDHGQYVRQD
                  280      290      300      310      320      330

    orf81.pep                      150      160      170      180      190      200
    IYNQGTVPQDSYLVPLVLVSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
15  orf81a       IYNQGTVPQDSYLVPLVLVSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG
                  340      350      360      370      380      390

    orf81.pep                      210      220      230
    CREGSVTGNLITGDAGSLNIRDGKAEYVYPQX
    |||||||||||||||||||||||||||||||||||
    orf81a       CREGSVTGNLITGDAGSLNIRDGKAEYVYPQX
                  400      410      420

```

The complete length ORF81a nucleotide sequence <SEQ ID 307> is:

25	1	ATGAAAAAAT	CCCTTTTCGT	TCTCTTTCTG	TATTCGTCCC	TACTTACTGC
	51	CAGCGAAATT	GCTTATCGCT	TTGTATTCCG	AATTGAAACC	TTACCGGCTG
	101	CAAAAATGGC	AGAAACGTTT	GCGCTGACAT	TTGTGATTGC	TGCGCTGTAT
	151	CTGTTTGCGC	GTTATAAGGC	AACGCGTTTG	TTGATTGCGG	TGTTTTTCGC
	201	GTTCAGCATT	ATTGCCAACA	ATGTGCATTA	CTCGCTTTAT	CAAAAGCTGGA
30	251	TAACGGGCAT	TAATTATTGG	CTGATGCTGA	AAGAGATTAC	CGAAGTTGGC
	301	GGCGCAGGGG	CGTCGATGTT	GGATAAGTTG	TGGCTGCCTG	CGTTGTGGGG
	351	CGTGTTGGAA	GTCATGTTGT	TTTGCAGCCT	TGCCAAGTTT	CGCCGTAAGA
	401	CGCATTTTTT	TGCCGATATA	CTGTTTGCCT	TCCTAATGCT	GATGATTTTT
	451	GTGCGTTCGT	TCCGACGAA	ACAAGAACAC	GGTATTTTCG	CCAAACCCGAC
35	501	ATACAGCCGC	ATCAAAGCCA	ATTATTTTCA	CTTCGGTTAT	TTTGTCCGAC
	551	GCGTGTGGCC	GTATCAGTTG	TTTGATTTAA	GCAAGATTCC	TGTGTTCAAA
	601	CAGCCTGCCT	CAAGCAGAAT	CGGGCAAGGC	AGTATTCAAA	ATATCGTCCT
	651	GATTATGGGC	GAAAGCGAAA	GCGCGGCGCA	TTTGAAATTG	TTTGGCTACG
	701	GGCGCGAAAC	TTGCGCGTTT	TTGACCCAGC	TTTCGCAAGC	CGATTTTAA
40	751	CCGATTGTGA	AACAAAGTTA	TTCCGCAAGC	TTTATGACGG	CAGTATCCCT
	801	GCCCAGTTTC	TTTAACGTCA	TACCGCATGC	CAACGGCTTG	GAACAAATCA
	851	GCGGCGGCGA	TATTGTGGAT	AAGTACGACA	ACACCATCCA	CAAAACCGAC
	901	CAAATGATTC	AAACCGTATT	CGAGCAGCTG	CAAAAGCAGC	CTGACGGCAA
	951	CTGGCTGTTT	GCCTATACCT	CCGATCATGG	CCAGTATGTT	GCCCAAGATA
45	1001	TCTACAATGA	AGGCAAGGTG	CAGCCCGACA	GCTATCTCGT	GCCGCTGGTG
	1051	TTGTACAGCC	CGGATAAGGC	CGTGCAACAG	GCTGCCAACC	AGGCTTTTGC
	1101	GCCTTGCGAG	ATTGCCTTCC	ATCAGCAGCT	TTCAACGTTT	CTGATTCACA
	1151	CGTTGGGCTA	CGATATGCCG	GTTTCAGGTT	GTCCGGAAGG	CTCGGTAACG
	1201	GGCAACCGTA	TTACGGGTGA	TGCAGGCAGC	TTGAACATTG	GCGACGGCAA
	1251	GGCGGAATAT	GTTTATCCGC	AATGA		

50 This encodes a protein having amino acid sequence <SEQ ID 308>:

55

1	MKKSFLVLFL	YSSLLTASEI	AYRFVFGIET	LPAAKMAETF	ALTFFVIAALY
51	<u>LFARYKATRL</u>	<u>LIAVFFAFSI</u>	<u>IANNVHYAVY</u>	QSWITGINYW	LMKEITEVVG
101	GAGASMLDKL	WLPALWGLE	VMLFCSLAKF	RRKTHFSADI	LFAFLMLMIF
151	<u>VRSFDTKQEH</u>	<u>GISPKPTYSR</u>	<u>IKANYFSFGY</u>	<u>FVGRVLPHYQL</u>	<u>FDLXSLPVFK</u>
201	QPAPSRIGQG	SIQNIVLIMG	ESESAHLKL	FGYGRETSPF	LTQLSQADFK
251	PIVKQSYSAG	FMTAVSLPSF	FNVIPHANGL	EQISGGDIVD	KYDNTIHKTD
301	QMIQTVEEQL	QKQPDGNWLF	AYTSDHGQYV	RQDIYNQGTV	QPD SYLVPLV
351	LYSPDKAVQQ	AAQAFAFCE	IAFHQQLS TF	LIHTLGYDMP	VSGCREGSVT
401	GNLITGDAGS	LNIRDGKAEY	VYPQ*		

60 ORF81a and ORF81-1 show 77.9% identity in 524 aa overlap:

```

              10      20      30      40      50      60
orf81a.pep    MKKSLFVLFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFVIAALYLFARYKATRL
               ||||::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
65 orf81-1     MKKSFLTLVLYSSLLTASEIAYRFVFGIETLPAAKIAETFALTFVIAALYLFARYKVTRL
               10      20      30      40      50      60

```

-207-

		70	80	90	100	110	120
	orf81a.pep	LI	AVFFAFSII	ANNVHYAVYQ	SWITGINY	WMLKEITEV	GGAGASMLDKLWLPALWGVLE
5	orf81-1	LI	AVFFAFSII	ANNVHYAVYQ	SWMTGINY	WMLKEVTEV	GSAGASMLDKLWLPVWGVLE
		70	80	90	100	110	120
	orf81a.pep	130	140	150	160	170	180
10	orf81a.pep	VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTY	SRIKANYFSFGY				
	orf81-1	VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTY	SRIKANYFSFGY				
		130	140	150	160	170	180
	orf81a.pep	190	200	210	220	230	240
15	orf81a.pep	FVGRVLPYQLFDLSKIPVFKQPAPSRIGQGSIGNIVLIMGESESA	AHLKLF	GYGRETSPF			
	orf81-1	FVGRVLPYQLFDLSRIPAFKQPAPSKIGQGSVQIVLIMGESESA	AHLKLF	GYGRETSPF			
		190	200	210	220	230	240
20	orf81a.pep	250	260	270	280		
	orf81a.pep	LTQLSQADFKPIVKQSYSAGFMTAVSLPSFFNVI	PHANGLEQISGGD				
	orf81-1	LTRL	SQADFKPIVKQSYSAGFMTAVSLPSFFNAI	PHANGLEQISGGDTNMFRLAKEQGYE			
25		250	260	270	280	290	300
	orf81a.pep	-----					
30	orf81-1	TYFYSAQAENEMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQQGKHF					
		310	320	330	340	350	360
	orf81a.pep	-----		290	300	310	320
35	orf81-1	IVLHQ	RGSHPY	GALLQPD	KVFGEADIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLF		
		370	380	390	400	410	420
	orf81a.pep	330	340	350	360	370	380
40	orf81a.pep	AYTSDHGQYVRQDIYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
	orf81-1	AYTSDHGQYVRQDIYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
		430	440	450	460	470	480
45	orf81a.pep	390	400	410	420		
	orf81a.pep	LIHTLGYDMPVSGCREG	SVTGNLITGDAGSLNIRDGKA	EYVYPQX			
	orf81-1	LIHTLGYDMPVSGCREG	SVTGNLITGDAGSLNIRDGKA	EYVYPQX			
		490	500	510	520		

# 50 Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF81 and a predicted ORF (ORF81.ng) from *N. gonorrhoeae* of the N- and C-termini show 82.4 % and 97.5% identity in 85 and 121 overlap, respectively:

	orf81.pep	MKKSFLTLVLYSSLLTASEIAYPLELGIETLPAAKIAETFALT	FVIAALYL	FARNKVTRL	60
55	orf81ng	MKKSFLVLFYSSLLTASEIAYRFVFGIETLPAAKMAETFALT	FMIAALYL	FARYKASRL	60
	orf81.pep	LI	AVFFAFSII	ANNVHYADYQSWMT	85
60	orf81ng	LI	AVFFAFSII	ANNVHYAVYQSWMT	120
	orf81.pep			//	
	orf81ng			QTVFEQLQKTPDGNWLFAYTSDHGQYVRQD	433
	orf81ng	ALLQ	PQDKVFGEADIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLFAYTSDHGQYVRQD		433
65	orf81.pep	IYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG			493
	orf81ng	IYNQGTVPDSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG			493

-208-

```

orf81.pep    CREGSVTGNLITGDAGSLNIRDGKAEYVYPQ  524
             |||||:|||||
orf81ng      CREGSVTGNLITGDAGSLNIRNGKAEYVYPQ  524

```

The complete length ORF81ng nucleotide sequence <SEQ ID 309> is:

```

5      1  ATGAAAAAAT CCCTTTTCGT TCTCTTCTG TATTATCCC TACTTACCGC
      51  CAGCGAAATC GCCTATCGCT TTGTATTCGG AATTGAAACC TTACCGGCTG
     101  CAAAATGGC  GGAAACGTTT GCGCTGACAT TTATGATTGC TGCGCTGTAT
     151  CTGTTTGGCG GTTATAAGGC TTCGCGGCTG CTGATTGCGG TGTTTTTCGG
     201  GTTCAGCATG ATTGCCAACA ATGTGCATTA CGCGGTTTAT CAAAGCTGGA
    10   251  TGACGGGTAT  TAACTATTGG CTGATGCTGA AAGAGGTTAC CGAAGTCGGC
     301  AGCGCGGGCG CGTCGATGTT GGATAAGTTG TGGCTGCCTG CTTTGTGGGG
     351  CGTGCGGGAA GTCATGTTGT TTTGCAGCCT TGCCAAGTTC CGCCGTAAGA
     401  CGCATTTTTT TCCCGATATA CTGTTTGCCT TCCTAATGCT GATGATTTTC
     451  GTGCGTTCGT TCGACACGAA ACAAGAGCAC GGTATTTTCG CCAAACCGAC
    15   501  ATACAGCCGC  ATCAAAGCCA ATTATTTTCT CTTGCGTTAT TTTGTCGGGC
     551  GCGTGTGGCC GTATCAGTTG TTTGATTTAA GCAAGATCCC TGTGTTCAAA
     601  CAGCTGCTC  CAGCAAAAT CGGGCAAGGC AGTATTCAAA ATATCGTCCT
     651  GATTATGGGC  GAAAGCGAAA GCGCGGCGCA TTTGAAATTG TTTGGTTACG
     701  GGCGCGAAAC TTCGCCGTTT TTAACCCGGC TGTGCGCAAG CGATTTTAAG
    20   751  CCGATTGTGA  AACAAAGTTA TTCCGCAGGC TTTATGACGG CAGTATCCCT
     801  GCCCAGTTTC TTTAACGTCA TACCGCACGC CAACGGCTTG GAACAAATCA
     851  GCGCGGCGCA TACCAATATG TTCCGCCTCG CCAAAGAGCA GGGCTATGAA
     901  ACGTATTTTT ACAGTGCCCA GGCTGAAAAC CAAATGGCAA TTTTGAACCT
     951  AATCGGTAAG  AAATGGATAG ACCATCTGAT TCAGCCGACG CAACTTTAAG
    25   1001  ACGCAACCGG CGACAATATG CCCGATGAGA AGCTGCTGCC GTTGTTCGAC
     1051  AAAATCAATT TGCAGCAGGG CAGGCATTTT ATCGTGTGTC ACCAACGCGG
     1101  TTCCGACGCC CCATACGGCG CATTGTTGCA GCCTCAAGAT AAAGTATTCG
     1151  GCGAAGCCGA TATTGTGGAT AAGTACGACA ACACCATCCA CAAAACCGAC
     1201  CAAATGATT  AAACGTATT  CGAGCAGCTG CAAAAGCAGC CTGACGGCAA
    30   1251  CTGGCTGTTT  GCCTATACCT CCGATCATGG CCAGTATGTG CGCCAAGATA
     1301  TCTACAATCA AGGCACGGTG CAGCCCGACA GCTATATTGT GCCTCTGGTT
     1351  TTGTACAGCC CGGATAAGGC CGTGCAACAG GCTGCCAACC AGGCTTTTGC
     1401  GCCTTGCGAG ATTGCCTTCC ATCAGCAGCT TTCAACGTT  CTGATTCACA
     1451  CGTTGGGCTA CGATATGCCG GTTTCAGGTT GTCGCGAAGG CTCGGTAACA
    35   1501  GGCAACCTGA  TTACGGGCGA TGCAGGCAGC TTGAACATTC GCAACGGCAA
     1551  GGCGGAATAT GTTTATCCGC AATAA

```

This encodes a protein having amino acid sequence <SEQ ID 310>:

```

      1  MKKSLFVLFL YSSLLTASEI AYRFVFGIET LPAAKMAETF ALTFMIAALY
     51  LFARYKASRL LIAVFFAFSM IANNVHYAVY QSWMTGINYW LMLKEVTEVG
    40   101  SAGASMLDKL WLPALWGVAE VMLFCSLAKF RRKTHFSADI LFAFLMLMIF
     151  VRSEDTKQEH GISPKPTYSR IKANYFSFGY FVGRVLPYQL FDLSKIPVEK
     201  QPAPSKIGQG SIQNIVLIMG ESESAHLKL FGYGRETSPP LTRLQADFK
     251  PIVKQSYSAG FMTAVSLPSF FNVIPHANGL EQISGGDTNM FRLAKEQGYE
     301  TYFYSAQAE  NMAILNLIGK KWIDHLIPT QLGYNNGDNM PDEKLLPLFD
    45   351  KINLQQGRHF IVLHQGRSHA PYGALLQPQD KVFGEADIVD KYDNTIHKTD
     401  QMIQTVFEQL QKQPDGNWLF AYTSDHGQYV RQDIYNQGTV QPDSYIVPLV
     451  LYSYDKAVQQ AANQAFAPCE IAFHQQLSTF LIHTLGIDMP VSGCREGSVT
     501  GNLTGDAGS LNIRNGKAEY VYPQ*

```

ORF81ng and ORF81-1 show 96.4% identity in 524 aa overlap:

```

50      10      20      30      40      50      60
orf81ng-1.pep MKKSLFVLFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFMIAALYLFARYKASRL
             |||||:|||||
orf81-1        MKKSFLTLYSSLLTASEIAYRFVFGIETLPAAKIAETFALTFVIAALYLFARYKVTRL
             |||||:|||||
    55      70      80      90      100     110     120
orf81ng-1.pep LIAVFFAFSMIANNVHYAVYQSWMTGINYWMLKEVTEVGSAGASMLDKLWLPALWGVAE
             |||||:|||||
orf81-1        LIAVFFAFSIIANNVHYAVYQSWMTGINYWMLKEVTEVGSAGASMLDKLWLPVWLVLE
             |||||:|||||
    60      70      80      90      100     110     120
orf81ng-1.pep VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRIKANYFSFGY
             |||||:|||||
orf81-1        VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRIKANYFSFGY
             |||||:|||||
    65      130     140     150     160     170     180
orf81ng-1.pep VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRIKANYFSFGY
             |||||:|||||
orf81-1        VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRIKANYFSFGY
             |||||:|||||

```

-209-

		130	140	150	160	170	180
5	orf81ng-1.pep	190	200	210	220	230	240
	orf81-1	190	200	210	220	230	240
10	orf81ng-1.pep	250	260	270	280	290	300
	orf81-1	250	260	270	280	290	300
15	orf81ng-1.pep	310	320	330	340	350	360
	orf81-1	310	320	330	340	350	360
20	orf81ng-1.pep	370	380	390	400	410	420
	orf81-1	370	380	390	400	410	420
25	orf81ng-1.pep	430	440	450	460	470	480
	orf81-1	430	440	450	460	470	480
30	orf81ng-1.pep	490	500	510	520		
	orf81-1	490	500	510	520		

Furthermore, ORF81ng shows significant homology to an *E.coli* OMP:

40	gi 1256380 (U50906) outer membrane adherence protein-associated protein [E. coli] Length = 547 Score = 87.4 bits (213), Expect = 2e-16 Identities = 122/468 (26%), Positives = 198/468 (42%), Gaps = 70/468 (14%)
45	Query: 25 VFGIETLPAAKMAETFA-LTFMIAALYLFAFYKAS--RLLIAVFFAFSMIANNVHYAVYQ 81 VFGI L A+ A L F + + + R + RLL+A F + A ++ ++Y Sbjct: 29 VFGITNLVASSGAHMQRLFFVLTLVVKRISSLPLRLLVAAPFVL-LTAADMSISLY- 86
50	Query: 82 SWMT-----GINYWMLKEVTEVGSAGASMLDKLWLPALWGVAEVMLFCSLAKFRRKT 134 SW T G ++ + EV A ML ++ P L A + L + Sbjct: 87 SWCTFGTTFNDGFAISVLQSDPDEV---AKMLG-MYSPYLCAFAFLSLLFLAVIIKYDV 141
55	Query: 135 HFSADILFAFLMLMIFVRSF-----DTKQEHGISPKPTYSRKAN--YFSFGYFVG 183 + L+L++ S D K ++ SP SR +F+ YF Sbjct: 142 SLPTKKVTGILLLLIVISGSLFSACQFAYKDAKNKNAFSPYILASRFATYTPFFNLNYFAL 201
60	Query: 184 RVLPHYQ--LFDLSKIPVFKQPAPSKIGQSGSIQNIVLIMGESESAHLKLFYGYGRETSPFL 241 +Q L + +P F+ + I VLI+GES ++ L+GY R T+P + Sbjct: 202 AAKEHQRLLSIANTVPYFQL----SVRDTGIDTYVLIVGESVRVDNMSLYGYTRSTTPQV 257
65	Query: 242 TRLSQADFKPIVKQSYSAGFMTAVSLP---SFFNVIPHANGLEQISGGDTNMFRLAKEQG 298 +Q + Q+ S TA+S+P + +V+ H I N+ +A + G Sbjct: 258 E--AQRKQIKLFNQAISGAPYTALSVPLSLTADSVLSH-----DIHNPDPNIINMANQAG 310
70	Query: 299 YETYFYSAQA---ENQMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQ 355 ++T++ S+Q+ +N A+ ++ ++ + Y G DE LLP + Q Sbjct: 311 FQTFWLSSQSAFRQNGTAVTSI-----AMRAMETVYVRGF---DELLPLHLSQALQQ 359
	Query: 356 --QGRHFIVLHQRGSHAPYGALLQPQDKVFEADIVDK-YDNTIHKTDQMIQTVFEQLQK 412 Q + IVLH GSH P + VF D D YDN+IH TD ++ VFE L+ Sbjct: 360 NTQQKKLIVLHLNGSHEPACSAYPQSSAVFQPQDDQACDYDNSIHYTDSLLGQVFELLK- 418

Query: 413 QPDGNWLFAYTSDHG---QYVRQDIYNQG--TVQPDSYIVPL-VLYSP 454  
                   D          Y +DHG          ++++Y G          +Y VP+ + YSP  
 Sbjct: 419 --DRRASVMYFADHGLERDPTKKNVYFHGGREASQQAYHVPFIWYSP 464

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 37

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 311>:

```

1   ...ACCCTGCTCC TCTTCATCCC CCTCGTCCTC ACAC.GTGCG GCACACTGAC
51  CGGCATACTC GCCCaCGGCG GCGGCAAACG CTTTGCCGTC GAACAAGAAC
101 TCGTCGCCGC ATCGTCCCGC GCCGCCGTCA AAGAAATGGA TTTGTCCGCC
151 yTAAAGGAGC GCAAAGCCGC CyTTTACGTC TCCGTTATGG GCGACCAAGG
201 TTCGGGCAAC ATAAGCGGCG GACGCTACTC TATCGACGCA CTGATACGCG
251 GCGGCTACCA CAACAACCCC GAAAGTGCCA CCCAATACAG CTACCCCGCC
301 TACGACACTA CCGCCACCAC CAAATCCGAC GCGCTCTCCA GCGTAACCAC
351 TTCCACATCG CTTTGAACG CCCCCGCCGC CGyCyTGACG AAAAACAGCG
401 GACGCAAAGG CGAACGcTCC GCCGGACTGT CCGTCAACGG CACGGGCGAC
451 TACCGCAACG AAACCCTGCT CGCCAACCCC CGCGACGTTT CCTTCCTGAC
501 CAACCTCATC CAAACCGTCT TCTACCTGCG CGGCATCGAA GTCgTACCGC
551 CCGrATACGC CGACACCGAC GTATTCTGTA CCGTCGACGT A...
```

This corresponds to the amino acid sequence <SEQ ID 312; ORF83>:

```

1   ..TLLLFIPVL TXCGTLTGIL AHGGGKRFAV EQELVAASSR AAVKEMDLA
51  LKGRKAAXYV SVMGDQSGN ISGGRYSIDA LIRGGYHNNP ESATQYSYPA
101 YDTTATTKSD ALSSVTTS LLNAPAAHLT KNSGRKGRS AGLSVNGTGD
151 YRNETLLANP RDVSFLTNLI QTVFYLRGIE VVPPXYADTD VFTVTDV..
```

Further work revealed the complete nucleotide sequence <SEQ ID 313>:

```

1   ATGAAAACCC TGCTCCTCCT CATCCCCCTC GTCCTCACAG CCTGCGGCAC
51  ACTGACCGGC ATACCCGCCC ACGGCGGCGG CAAACGCTTT GCCGTGAAC
101 AAGAACTCGT CGCCGCATCG TCCCGCGCCG CCGTCAAAGA AATGGATTG
151 TCCGCCCTAA AAGGACGCAA AGCCGCCCTT TACGTCTCCG TTATGGGCGA
201 CCAAGGTTCT GGCAACATAA GCGGCGGACG CTACTCTATC GACGCACTGA
251 TACGCGGCGG CTACCACAAAC AACCCTGAAA GTGCCACCCA ATACAGCTAC
301 CCGGCCTACG AACTACCGC CACCACCAA TCCGACGCGC TCTCCAGCGT
351 AACCCTTCC ACATCGCTTT TGAACGCCCC CGCCGCGGCC CTGACGAAAA
401 ACAGCGGACG CAAAGGCGAA CGCTCCGCGG GACTGTCCGT CAACGGCAGC
451 GGCGACTACC GCAACGAAAC CCTGCTCGCC AACCCTCGCG ACGTTTCCTT
501 CCTGACCAAC CTCATCCAAA CCGTCTTCTA CCTGCGCGGC ATCGAAGTCG
551 TACCGCCCGA ATACGCCGAC ACCGACGTAT TCGTAACCGT CGACGTATTC
601 GGCACCGTCC GCAGCCGTAC CGAACTGCAC CTCTACAACG CCGAAACCTT
651 TAAAGCCCAA ACCAAGCTCG AATATTTCGC CGTTGACCGC GACAGCCGGA
701 AACTGCTGAT TACCCCTAAA ACCGCCGCTT ACGAATCCCA ATACCAAGAA
751 CAATACGCCC TTTGGACCGG CCCTTACAAA GTCAGCAAAA CCGTCAAAGC
801 CTCAGACCGC CTGATGGTCG ATTTCTCCGA CATTACCCCC TACGGCGACA
851 CAACCGCCCA AAACCGTCCC GACTTCAAAC AAAACAACGG TAAAAAACCC
901 GATGTCGGCA ACGAAGTCAT CCGCCGCGCG AAAGGAGGAT AA
```

This corresponds to the amino acid sequence <SEQ ID 314; ORF83-1>:

```

1   MKTLLLLIPL VLTACGTLTG IPAHGGGKRF AVEQELVAAS SRAAVKEMDL
51  SALKGRKAAL YVSVMGDQGS GNISGGRYSI DALIRGGYHN NPESATQYSY
101 PAYDTTATTK SDALSSVTTS TSLLNAPAAA LTKNSGRKGE RSAGLSVNGT
151 GDYRNETLLA NPDVSFLTN LIQTVFYLRG IEVVPPEYAD TDVFTVTDVF
201 GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLITPK TAAYESQYQE
251 QYALWTGPYK VSKTVKASDR LMVDFSDITP YGDTTAQNRD DFKQNNKGKP
```



301 DVGNEVIRRR KGG\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF83 shows 96.4% identity over a 197aa overlap with an ORF (ORF83a) from strain A of *N.*

5 *meningitidis*:

		10	20	30	40	50
orf83.pep		TLLLFIPVLTXCGTLTGILAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAX				
orf83a		MKTLLXLIPLVLTACGTLTGIPAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAL				
10		10	20	30	40	50
	60	70	80	90	100	110
orf83.pep	YVSVMGDQSGNISGGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTS					
15	orf83a	YVSVMGDQSGNISGGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTS				
		70	80	90	100	110
		120	130	140	150	160
20	orf83.pep	TSLLNAPAAXLTKNSGRKGRSAGLSVNGTGDYRNETLLANPRDVSFLTNIQTIFYLRG				
	orf83a	TSLLNAPAAALTKNSGRKGRSAGLSVNGTGDYRNETLLANPRDVSFLTNIQTIFYLRG				
		130	140	150	160	170
		180	190			
25	orf83.pep	IEVVPXYADTDVFVTVDV				
	orf83a	IEVVPPEYADTDVFVTVDFVGTVRSRTELHLYNAETLKAQTKLEYFAVDRDRSKLLIAPK				
		190	200	210	220	230
		240				

The complete length ORF83a nucleotide sequence <SEQ ID 315> is:

30	1	ATGAAAACCC	TGCTCNCCT	CATCCCCCTC	GTCCTCACAG	CCTGCGGCAC
	51	ACTGACCGGC	ATACCCGCCC	ACGGCGGCGG	CAAACGCTTT	GCCGTCGAAC
	101	AAGAACTCGT	CGCCGCATCG	TCCCGCGCCG	CCGTCAAAGA	AATGGACTTG
	151	TCCGCCCTGA	AAGGACGCAA	AGCCGCCCTT	TACGTCTCCG	TTATGGGCGA
	201	CCAAGGTTCG	GGCAACATAA	GCGGCGGACG	CTACTCTATC	GACGCACTGA
35	251	TACGCGGCGG	CTACCACAAC	AACCCCGAAA	GTGCCACCCA	ATACAGCTAC
	301	CCCGCCTACG	ACACTACCGC	CACCACCAAA	TCCGACGCGC	TCTCCAGCGT
	351	AACCACTTCC	ACATCGCTTT	TGAACGCCCC	CGCCGCCGCC	CTGACGAAAA
	401	ACAGCGGACG	CAAAGCGGAA	CGCTCCGCGG	GACTGTCCGT	CAACGGCACG
	451	GGCGACTACC	GCAACGAAAC	CCTGCTCGCC	AACCCCGCGG	ACGTTTCCTT
40	501	CCTGACCAAC	CTCATCCAAA	CCGTCTTCTA	CCTGCGCGGC	ATCGAAGTCG
	551	TACCGCCCGA	ATACGCGGAC	ACCGACGTAT	TCGTAACCGT	CGACGTATTC
	601	GGCACCGTCC	GCAGCCGCAC	CGAACTGCAC	CTCTACAACG	CCGAAACCCT
	651	TAAAGCCCAA	ACCAAGCTCG	AATATTTTCG	CGTTGACCGC	GACAGCCGGA
	701	AACTGCTGAT	TGCCCCTAAA	ACCGCGCCCT	ACGAATCCCA	ATACCAAGAA
45	751	CAATACGCCC	TCTGGATGGG	ACCTTACAGC	GTCGGCAAAA	CCGTCAAAGC
	801	CTCAGACCGC	CTGATGGTCG	ATTTCTCCGA	CATCACCCCC	TACGGCGACA
	851	CAACCGCCCA	AAACCGTCCC	GACTTCAAAC	AAAACAACGG	TAAAAAACCC
	901	GATGTCGGCA	ACGAAGTCAT	CCGCCGCCGC	AAAGGAGGAT	AA

This encodes a protein having amino acid sequence <SEQ ID 316>:

50	1	MKTLLXLIPL	VLTACGTLTG	IPAHGGGKRF	AVEQELVAAS	SRAAVKEMDL
	51	SALKGRKAAL	YVSVMGDQGS	GNISGGGRYSI	DALIRGGYHN	NPESATQYSY
	101	PAYDTTATTK	SDALSSVTTS	TSLLNAPAAA	LTKNSGRKGE	RSAGLSVNGT
	151	GDYRNETLLA	NPRDVSFLTNI	LIQTVFYLRG	IEVVPPEYAD	TDVFVTVDFV
	201	GTVRSRTELH	LYNAETLKAQ	TKLEYFAVDR	DSRKLLIAPK	TAAYESQYQE
55	251	QYALWMGPYS	VGKTVKASDR	LMVDFSDITP	YGDTTAQNRP	DFKQNGKKP
	301	DVGNEVIRRR	KGG*			

ORF83a and ORF83-1 show 98.4% identity in 313 aa overlap:

	10	20	30	40	50	60
orf83a.pep	MKTLLXLIPLVLTACGTLTGIPAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAL					

10

15

20

25

30

35 Homology with a predicted ORF from *N.gonorrhoeae*

ORF83 shows 94.9% identity over a 197aa overlap with a predicted ORF (ORF83.ng) from *N. gonorrhoeae*:

40

45

50

The complete length ORF83ng nucleotide sequence <SEO ID 317> is:

55

60

65

-213-

5  
 651 TAAAGCCCAA ACCAAGCTCG AATATTTTCGC CGTCGACCGC GACAGCCGGA  
 701 AACTGCTGAT TGCCCTAAA ACCGCCGCT ACGAATCCCA ATACCAAGAA  
 751 CAATACGCCC TCTGGATGGG ACCTTACAGC GTCGGCAAAA CCGTCAAAGC  
 801 CTCAGACCGC CTGATGGTCG ATTTCTCCGA CATCACCCC TACGGCGACA  
 851 CAACCGCCA AAACCGTCCC GACTTCAAAC AAAACAACGG TAAAAACCCC  
 901 GATGTCGGCA ACGAAGTCAT CCGCCGCCGC AAAGGAGGAT AA

This encodes a protein having amino acid sequence <SEQ ID 318>:

10  
 1 MKTLLLLLIPL VLTACGTLTG IPAHHGGKRF AVEQELVAAS SRAAVKEMDL  
 51 SALKGRKAAL YVSMGDQGS GNISGGRYSI DALIRGGYHN NPDSATRYSY  
 101 PAYDTTATTK SDALSGVTTs TSLNAPAAA LTKNNGRKGE RSAGLSVNGT  
 151 GDYRNETLLA NPRDVSFLTn LIQTVFYLRG IEVVPPEYAD TDVFVTVDFV  
 201 GTVRSRTELH LYNATLKAQ TKLEYFAVDR DSRKLLIAPK TAAYESQYQE  
 251 QYALWMGPYS VGKTVKASDR LMVDFSDITP YGDTTAQNRP DFKQNNNGKNP  
 301 DVGNEVIRRR KGG\*

15 ORF83ng and ORF83-1 show 97.1% identity in 313 aa overlap

		10	20	30	40	50	60
20	orf83-1.pep	MKTLLLLLIPLVLTACGTLTGIPAHHGGKRF	AVEQELVAASSRAAVKEMDL	SALKGRKAAL			
	orf83ng	MKTLLLLLIPLVLTACGTLTGIPAHHGGKRF	AVEQELVAASSRAAVKEMDL	SALKGRKAAL			
		10	20	30	40	50	60
25	orf83-1.pep	YVSMGDQGS	GNISGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTs				
	orf83ng	YVSMGDQGS	GNISGGRYSIDALIRGGYHNNPDSATRYSYPAYDTTATTKSDALSGVTTs				
		70	80	90	100	110	120
30	orf83-1.pep	TSLNAPAAA	LTKNNGRKGE	RSAGLSVNGTGDYRNETLLANPRDVSFLTnLIQTVFYLRG			
	orf83ng	TSLNAPAAA	LTKNNGRKGE	RSAGLSVNGTGDYRNETLLANPRDVSFLTnLIQTVFYLRG			
		130	140	150	160	170	180
35	orf83-1.pep	IEVVPPEYADTDVFVTVDFVFGTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLITPK					
	orf83ng	IEVVPPEYADTDVFVTVDFVFGTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLITPK					
		190	200	210	220	230	240
40	orf83-1.pep	TAAYESQYQE	QYALWGPYS	<u>VGKTVKASDR</u>	LMVDFSDITPYGDTTAQNRPDFKQNNNGKGP		
	orf83ng	TAAYESQYQE	QYALWGPYS	<u>VGKTVKASDR</u>	LMVDFSDITPYGDTTAQNRPDFKQNNNGKNP		
		250	260	270	280	290	300
45	orf83-1.pep	DVGNEVIRRR	KGGX				
	orf83ng	DVGNEVIRRR	KGGX				
50		310					

Based on this analysis, including the presence of a putative ATP/GTP-binding site motif A (P-loop) in the gonococcal protein (double-underlined) and a putative prokaryotic membrane lipoprotein lipid attachment site (single-underlined), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 38**

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 319>:

```

      1  ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAAACATT
5      51  AAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
      101  AAGCCATACG CCGTAAAGTA TTTACGAACA TAAAAGGCTT GAAAATACCG
      151  CACACCTACA TAGAAACGGA CGCAAAAAAG CTGCCGAAAT CGACAGATGA
      201  GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
10     251  TCGGGTCTAT TGTCATTGTA GATGAAGCTC AAGACGTATG GCCGGCACGC
      301  TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
      351  ACATCAGGGC ATTGATATAT TTGTTTGTAC TCAAGGTCCT AAGCTTCTAG
      401  ATCAAAATCT TAGAACGCTT GTACGGAAAC ATTACCACAT CGCTTCAAAC
      451  AAGATGGGTA TCGGTACGCT TTTAGAATGG AAAATATGCG CGGACGATCC
15     501  CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
      551  AAGTTTATGA CTTGTATsrr TmmGCGGAAG TTCATACCGT AAATAAGGTC
      601  AAGCGGTCAA AGTGGTTTTA CACTCTGCCa GTAATAGTAT TGCTGATTCC
      651  CGTGTTTGTC GGCCTGTCTT ATAAAATGTT GagCaGTTAC GGAAAAAAC
      701  aGGAAGAACC CGCAGCACA GAATCGGCGG CAACAGAACA GCAGGCAGTA
20     751  CTTCCGGATA AAACAGAAGG CGAGCCGGTA AATAACGGCA ACCTTACCGC
      801  AGATATGTTT GTTCCGACAT TGTCCGAaAA ACCCGrAAGC AAGCgaTTT
      851  ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTATA
      901  GAAGGCGGAA GAACCGGATG CGCCTGCTAT TCGCaTCAAG GGACGGCAt
      951  gaAAGAAGTG ACGGaGTTGA TGTGcgaAgG aCTATGTaAA AaAcGGCTTG
25    1001  CCGTTTAACC CaTACAAAGA AGAAAGCCAA GGCAGGAAG TTCAGCAAAG
      1051  CGCGCAgCAA CATTCGGACA GGGCGcCAAG TTGCCACATT GGGCGGAAAA
      1101  CCGTAGCAGA ACCTAATGTA CGATAATTGG GAAGAACGCG GGAAACCGTT
      1151  TGAAGGAATC GGaCGGGGGC GTGGTCGGAT CGGCAAAC TG A

```

This corresponds to the amino acid sequence <SEQ ID 320; ORF84>:

```

      1  MAEICLITGT PGSGKTLKMV SMMADEMFK PDEKAIRRKV FTNIKGLKIP
30     51  HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDVWPAR
      101  SAGSKIPENV QWLNTHRHQG IDIFVLTQGP KLLDQNLRTL VRKHYHIASN
      151  KGMGRTLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYX XAEVHTVNKV
      201  KRSKWFYTLF VIVLLIPVFV GLSYKMLSSY GKKQEEPAAG ESAATEQQAV
35     251  LPDKTEGEPV NNGNLTADMF VPTLSEKPKXS KPIYNGVRQV RTFEYIAGCI
      301  EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPFYKE ESQGEVQQS
      351  AQQHSRAQV ATLGKPKXQN LMYDNWEERG KPFEIGIGGV VGSAN*

```

Further work revealed the complete nucleotide sequence <SEQ ID 321>:

```

      1  ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAAACATT
40     51  AAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
      101  ACGGCATACG CCGTAAAGTA TTTACGAACA TAAAAGGCTT GAAAATACCG
      151  CACACCTACA TAGAAACGGA CGCAAAAAAG CTGCCGAAAT CGACAGATGA
      201  GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
      251  TCGGGTCTAT TGTCATTGTA GATGAAGCTC AAGACGTATG GCCGGCACGC
45     301  TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
      351  ACATCAGGGC ATTGATATAT TTGTTTGTAC TCAAGGTCCT AAGCTTCTAG
      401  ATCAAAATCT TAGAACGCTT GTACGGAAAC ATTACCACAT CGCTTCAAAC
      451  AAGATGGGTA TCGGTACGCT TTTAGAATGG AAAATATGCG CGGACGATCC
50     501  CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
      551  AAGTTTATGA CTTGTACGAA TCAGCGGAAG TTCATACCGT AAATAAGGTC
      601  AAGCGGTCAA AGTGGTTTTA CACTCTGCCa GTAATAGTAT TGCTGATTCC
      651  CGTGTTTGTC GGCCTGTCTT ATAAAATGTT GAGCAGTTAC GGAAAAAAC
      701  AGGAAGAACC CGCAGCACA GAATCGGCGG CAACAGAACA GCAGGCAGTA
      751  CTTCCGGATA AAACAGAAGG CGAGCCGGTA AATAACGGCA ACCTTACCGC
55     801  AGATATGTTT GTTCCGACAT TGTCCGAAAA ACCCGAAAGC AAGCCGATTT
      851  ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTATA
      901  GAAGGCGGAA GAACCGGATG CGCCTGCTAT TCGCaTCAAG GGACGGCATT
      951  GAAAGAAGTG ACGGAGTTGA TGTGCAAGGA CTATGTAAAA AACGGCTTGC
60    1001  CGTTTAACCC ATACAAAGAA GAAAGCCAAG GGCAGGAAGT TCAGCAAAGC
      1051  GCGCAGCAAC ATTTCGACAG GGCgCAAGTT GCCACATTGG GCGGAAAACC
      1101  GTAGCAGAAC CTAATGTACG ATAATTGGGA AGAACGCGGG AAACCGTTTG
      1151  AAGGAATCGG CGGGGGCGTG GTCGGATCGG CAAACTGA

```

This corresponds to the amino acid sequence <SEQ ID 322; ORF84-1>:

-215-

1 MAEICLITGT PGSGKTLKVM SMMANDEMFK PDENGIRRKV FTNIKGLKIP  
 51 HTYIETDAKK LPKSTDEQLS AHDMEYEWIKK PENIGSIVIV DEAQDVWPAR  
 101 SAGSKIPENV QWLNTHRHQG IDIFVLTQGP KLLDQNLRTL VRKHYHIASN  
 151 KMGMRITLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYE SAEVHTVNKV  
 5 201 KRSKWFTLP VIVLLIPVFV GLSYKMLSSY GKKQEEPAAQ ESAATEQQAV  
 251 LPDKTEGEPV NNGNLTADMV VPTLSEKPES KPIYNGVRQV RTFEYIAGCI  
 301 EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQGQEVQQS  
 351 AQQHSRAQV ATLGKPK\*QN LMYDNWEERG KPFEIGGGV VGSAN\*

Computer analysis of this amino acid sequence gave the following results:

10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF84 shows 93.9% identity over a 395aa overlap with an ORF (ORF84a) from strain A of *N.*

*meningitidis*:

		10	20	30	40	50	60
15	orf84.pep	MAEICLITGT PGSGKTLKVM SMMANDEMFKPDEKAIRRKVFTNIKGLKIPHTYIETDAKK      :					
	orf84a	MAEICLITGT PGSGKTLKVM SMMANDEMFKPDENGIRRKVFTNIKGLKIPHTYIETDAKK 10 20 30 40 50 60					
		70	80	90	100	110	120
20	orf84.pep	LPKSTDEQLSAHDMEYEWIKK PENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG 					
	orf84a	LPKSTDEQLSAHDMEYEWIKK PENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG 70 80 90 100 110 120					
		130	140	150	160	170	180
25	orf84.pep	IDIFVLTQGP KLLDQNLRTLVRKHYHIASN KMGMRITLLEWKICADDPVKMASSAFSSIYT 					
	orf84a	IDIFVLTQGS KLLDQNLRTLVRKHYHIASN KMGMRITLLEWKICADDPVKMASSAFSSIYT 130 140 150 160 170 180					
		190	200	210	220	230	240
30	orf84.pep	LDKKVYDLYXXAEVHTVNKV KRSKWFTLPVIVLLIPVFVGLSYKMLSSYGKKQEEPAAQ 					
	orf84a	LDKKVYDLYESA EVHTVNKV KRSKWFTLPVILLIPVFVGLSYKMLSSYGKKQEEPAAQ 190 200 210 220 230 240					
		250	260	270	280	290	300
35	orf84.pep	ESAATEQQAVLPDKTEGEPV NNGNLTADMVPTLSEKPKSPIYNGVRQV RTFEYIAGCI      :					
	orf84a	ESAATEHQAVFDKTEGEPV NNGNLTADMVPTLSEKPKSPIYNGVRQV RTFEYIAGCV 250 260 270 280 290 300					
		310	320	330	340	350	360
40	orf84.pep	EGGRTGCACYSHQGTALKEV TELMCKDYVK NGLPFNPYKEESQGQEVQQSAQQHSDRAQV      :					
	orf84a	EGGRTGCTCYSHQGTALKEITKEMCKDYARNGLPFNPYKEESQGRDVQQSEQHSDRPQV 310 320 330 340 350 360					
		370	380	390			
45	orf84.pep	ATLGKPKXQNL MYDNWEERGKPFEGIGGGV VGSANX 					
	orf84a	ATLGKPKWQNL MYDNWQERGKPFEGIGGGV VGSANX 370 380 390					

The complete length ORF84a nucleotide sequence <SEQ ID 323> is:

55 1 ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAACATT  
 51 AAAAATGGTT TCCATGATGG CAAACGATGA AATGTTTAAG CCGGATGAAA  
 101 ACGGCATACG CCGTAAAGTA TTTACGAACA TCAAAGGCTT GAAGATACCG  
 151 CACACCTACA TAGAAACGGA CGCGAAAAAG CTGCCGAAAT CGACAGATGA  
 201 GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA  
 60 251 TCGGGTCTAT TGTCATTGTA GATGAAGCTC AAGACGTATG GCCGGCACGC  
 301 TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG  
 351 ACATCAGGCG ATTGATATAT TTGTTTGAC TCAAGGCTCT AAGCTTCTAG  
 401 ATCAAAATCT TAGAACGCTT GTACGGAAAC ATTACCACAT CGCTTCAAAC  
 451 AAGATGGGTA TCGGTACGCT TTTAGAAATGG AAAATATGCG CGGACGATCC

501 CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA  
 551 AAGTTTATGA CTGTGTACGAA TCAGCGGAAG TTCATACCGT AAATAAGGTC  
 601 AAGCGGTCAA AATGGTTTTA TACTCTGCCA GTAATAATAT TGCTGATTCC  
 651 CGTTTTTGTG GGCCTGTCTT ATAAAATGTT AAGTAGTTAT GGAAAAAAC  
 701 AGGAAGAACC CGCAGCACAA GAATCGGCGG CAACAGAACA TCAGGCAGTA  
 751 TTTCAGGATA AAACAGAAGG CGAGCCGGTA AACAACGGTA ACCTTACCGC  
 801 AGATATGTTT GTTCCGACAT TGTCCGAAAA ACCCGAAAGC AAGCCGATTT  
 851 ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTGTA  
 901 GAAGGCGGAA GAACCGGATG CACATGCTAT TCGCATCAAG GGACGGCATT  
 951 GAAAGAAATT ACAAAGGAAA TGTGCAAGGA TTACGCAAGA AACGGATTGC  
 1001 CGTTTAACCC ATATAAGAA GAAAGCCAAG GCGGGATGT CCAGCAAAGT  
 1051 GAGCAGCACC ATTCGGACAG ACCGCAAGTT GCCACGTTGG GCGGAAAGCC  
 1101 GTGCAAAAT CTTATGTATG ATAATTGGCA GGAGCGCGGA AAACCGTTTG  
 1151 AAGGAATCGG CGGGGGCGTG GTCGGATCGG CAAACTGA

15 This encodes a protein having amino acid sequence <SEQ ID 324>:

1 MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRVK FTNIKGLKIP  
 51 HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDVWPAR  
 101 SAGSKIPENV QWLNTHRHQG IDIFVLTQGS KLLDQNLRTL VRKHYHIASN  
 151 KMGMRITLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYE SAEVHTVNKV  
 201 KRSKWFYTLV VIILLIPVGV GLSYKMLSSY GKKQEPPAAQ ESAATEHQAV  
 251 FQDKTEGEPV NNGNLTADMF VPTLSEKPES KPIYNGVRQV RTFEYIAGCV  
 301 EGGRTGCTCY SHQGTALKEI TKEMCKDYAR NGLPFNPYKE ESQGRDVQQS  
 351 EQHHSRDPQV ATLGKWPQON LMYDNWQERG KPFEGIGGGV VGSAN\*

ORF84a and ORF84-1 show 95.2% identity in 395 aa overlap:

25		10	20	30	40	50	60
	orf84a.pep	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDENGIRRVKFTNIKGLKIPHTYIETDAKK					
	orf84-1	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDENGIRRVKFTNIKGLKIPHTYIETDAKK					
30		10	20	30	40	50	60
	orf84a.pep	LPKSTDEQLSAHDMYEWIKKPENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG					
	orf84-1	LPKSTDEQLSAHDMYEWIKKPENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG					
35		70	80	90	100	110	120
	orf84a.pep	IDIFVLTQGSKLLDQNLRTLVRKHYHIASNKMGMRITLLEWKICADDPVKMASSAFSSIYT					
	orf84-1	IDIFVLTQGSKLLDQNLRTLVRKHYHIASNKMGMRITLLEWKICADDPVKMASSAFSSIYT					
40		130	140	150	160	170	180
	orf84a.pep	LDKKVYDLYESADEVHTVNKVKRSKWFYTLVVIILLIPVGVGLSYKMLSSYGKKQEPPAAQ					
	orf84-1	LDKKVYDLYESADEVHTVNKVKRSKWFYTLVVIILLIPVGVGLSYKMLSSYGKKQEPPAAQ					
45		190	200	210	220	230	240
	orf84a.pep	ESAATEHQAVFQDKTEGEPVNNNGNLTADMFVPTLSEKPESKPIYNGVRQVRTFEYIAGCV					
	orf84-1	ESAATEHQAVFQDKTEGEPVNNNGNLTADMFVPTLSEKPESKPIYNGVRQVRTFEYIAGCI					
50		250	260	270	280	290	300
	orf84a.pep	EGGRTGCTCYSHQGTALKEITKEMCKDYARNGLPFNPYKEESQGRDVQQSSEQHHSRDPQV					
	orf84-1	EGGRTGCTCYSHQGTALKEITKEMCKDYARNGLPFNPYKEESQGRDVQQSSEQHHSRDPQV					
55		310	320	330	340	350	360
	orf84a.pep	ATLGKWPQONLMYDNWQERGKPFEGIGGGVVGSANX					
	orf84-1	ATLGKWPQONLMYDNWQERGKPFEGIGGGVVGSANX					
60		370	380	390			
	orf84a.pep	ATLGKWPQONLMYDNWQERGKPFEGIGGGVVGSANX					
	orf84-1	ATLGKWPQONLMYDNWQERGKPFEGIGGGVVGSANX					
65		370	380	390			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF84 shows 94.2% identity over a 395aa overlap with a predicted ORF (ORF84.ng) from *N. gonorrhoeae*:

5	orf84.pep	MAEICLITGTPGSGKTLKMSMMANDEMFKPDEKAIRRKVFTNIKGLKIPHTYIETDAKK	60
	orf84ng	MAEICLITGTPGSGKTLKMSMMANDEMFKPDENGVRKRVFTNIKGLKIPHTHIETDAKK	60
10	orf84.pep	LPKSTDEQLSAHDMYEWIKKPENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG	120
	orf84ng	LPKSTDEQLSAHDMYEWIKKPENVGAIIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG	120
15	orf84.pep	IDIFVLTQGPQLLDQNLRLTLVRKHYHIAANKMGMRTLLEWKICADDPVKMASSAFSSIYT	180
	orf84ng	IDIFVLTQGPQLLDQNLRLTLVRKHYHIAANKMGLRTLLEWKVCADDPVKMASSAFSSIYT	180
20	orf84.pep	LDKKVYDLYXXAEVHTVNKVKRSKWFTLPIVLLIPVFLVGLSYKMLSSYGKKQKEEPAAQ	240
	orf84ng	LDKKVYDLYESAEIHTVNKVKRSKWFTYALPVIILLIPLFVGLSYKMLGSYGKKQKEEPAAQ	240
25	orf84.pep	ESAATEQQAVLPDKTEGEPVNNGLTADMVPTLSEKPKSKPIYNGVRQVRTFEYIAGCI	300
	orf84ng	ESAATEQQAVLPDKTEGESVNNGLTADMVPTLPEKPESKPIYNGVRQVRTFEYIAGCI	300
30	orf84.pep	EGGRTGCACYSHQGTALKEVTELMCKDYVKNGLPFNPKYKEESQGQEVQSSAQQHSRAQV	360
	orf84ng	EGGRTGCTCYSHQGTALKEVTELMCKDYVKNGLPFNPKYKEESQGQEVQSSAQQHSRAQV	360
35	orf84.pep	ATLGGKPKQNLMYDNWEERGKPFEGIGGGVGSAN	395
	orf84ng	ATLGGKPKQNLMYDNWEERGKPFEGIGGGVGSAN	395

The complete length ORF84ng nucleotide sequence <SEQ ID 325> is:

1	ATGGCAGAAA	TCTGTTT	GAT	AACCGGCACG	CCCGGTT	CAG	GGAAAACATT
51	AAAAATGTT	TCCATGAT	G	CAACGATGA	AATGTTT	AAG	CCAGATGAAA
101	ACGGCGTACG	CCGTAAAGTA	TTTACGAACA	TCAAAGGTTT	GAAGATACCG		
151	CACACCCACA	TAGAAACAGA	CGCAAAGAAG	CTGCCGAAAT	CAACCGATGA		
201	ACAGCTTTCG	GCGCATGATA	TGTATGAATG	GATCAAGAAG	CCTGAAAacg		
251	tccgcgCAAT	CGTTATTGTC	GATGAGGCGC	AAGACGTATG	GCCCGCACGC		
301	TccgCAGGTT	CGAAAATCCC	CGAAAACGTC	CAATGGCTGA	ACACACACAG		
351	GCATCAGGGC	ATAGATATAT	TTGTATTGAC	ACAAGGTCCT	AAACTCTTAG		
401	ATCAGAACTT	GCGAAACATTG	GTAAAAAGAC	ATTACCACAT	TGCGGCCAAC		
451	AAAATGGGTT	TGCGTACCCT	GCTTGAATGG	AAAGTATGCG	CGGATGACCC		
501	GGTAAAAATG	GCATCAAGTG	CATTTTCCAG	TATCTACACA	CTGGATAAAA		
551	AAGTTTATGA	CTTGTACGAA	TCCGCAGAAA	TTACACCGGT	AAACAAAGTC		
601	AAGCGTTCAA	AATGGTTTTA	TGCATTGCCC	GTCATCATAT	TATTGATTCC		
651	GCTATTTGTC	GGTTTGTCTT	ACAAAATGTT	GGGCAGTTAC	GGAAAAAAC		
701	AGGAAGAACC	CGCAGCACAA	GAATCGGCGG	CAACAGAACA	GCAGGCAGTA		
751	CTTCCGGATA	AAACAGAAGG	AGAATCGGTG	AATAACGGAA	ACCTTACGGC		
801	AGATATGTTT	GTTCCGACAT	TGCCCGAAAA	ACCCGAAAGC	AAGCCGATTT		
851	ATAACGGTGT	AAGGCAGGTA	AGGACCTTTG	AATATATAGC	AGGCTGTATA		
901	GAAGGCGGAA	GAACCGGATG	CACCTGCTAT	TCGCATCAAG	GGACGGCATT		
951	GAAAGAAGTG	ACGGAGTTGA	TGTGCAAGGA	CTATGTAAAA	AACGGCTTGC		
1001	CGTTTAAACC	ATACAAAGAA	GAAAGCCAAG	GGCAGGAAGT	TCAGCAAAGC		
1051	GCGCAGCAAC	ATTCGACAG	GGCGCAAGTT	GCCACCTTGG	GCGGAAAACC		
1101	GCAGCAGAAC	CTAATGTACG	ACAATTGGGA	AGAACGCGGG	AAACCGTTTG		
1151	AAGGAATCGG	CGGGGGCGTG	GTCCGATCGG	CAACTGA			

This encodes a protein having amino acid sequence <SEQ ID 326>:

1	MAEICLITGT	PGSGKTLKMV	SMMANDEMFK	PDENGVRKRV	FTNIKGLKIP
51	HTHIETDAKK	LPKSTDEQLS	AHDMYEWIKK	PENVGAIVIV	DEAQDVWP
101	SAGSKIPENV	QWLNTHRHQG	IDIFVLTQGP	KLLDQNLRLT	VKRHYHIAAN
151	KMGLRTLLEW	KVCADDPVKM	ASSAFSSIYT	LDKKVYDLYE	SAEIHTVNKV
201	KRSKWYFALP	VIILLIPLFV	GLSYKMLGSY	GKKQKEEPAAQ	ESAATEQQAV
251	LPDKTEGESV	NNGLTADMV	VPTLPEKPES	KPIYNGVRQV	RTFEYIAGCI
301	EGGRTGCTCY	SHQGTALKEV	TELMCKDYVK	NGLPFNPKYKE	ESQGQEVQSS
351	AQQHSDRAQV	ATLGGKPKQN	LMYDNWEERG	KPFEGIGGGV	VGSAN*

ORF84ng and ORF84-1 show 95.4% identity in 395 aa overlap:

		10	20	30	40	50	60
	orf84-1.pep	MAEICLITGTPGSGKTLK	MVSMMA	NDEMFK	PDENGIR	KVFTNIK	GLKIPHTYIETDAKK
5	orf84ng	MAEICLITGTPGSGKTLK	MVSMMA	NDEMFK	PDENGVR	KVFTNIK	GLKIPHTHIETDAKK
		10	20	30	40	50	60
	orf84-1.pep	LPKSTDEQLSAHDMYEWIK	KPENIGS	IIVDEA	QDVWP	PARSAGS	KIPENVQWLNTHRHQG
10	orf84ng	LPKSTDEQLSAHDMYEWIK	KPENVGAI	VIVDEA	QDVWP	PARSAGS	KIPENVQWLNTHRHQG
		70	80	90	100	110	120
	orf84-1.pep	IDIFVLTQGP	KLLDQNL	RTRLV	RKHYH	IASNKM	GMRTLLEWKICADDPVKMASSAFSSIYT
15	orf84ng	IDIFVLTQGP	KLLDQNL	RTRLV	RKHYH	IAANKM	GLRTLLEWKVCADDPVKMASSAFSSIYT
		130	140	150	160	170	180
	orf84-1.pep	LDKKVYDLYESA	EIVTNK	VKR	SKW	FYTL	LPVIVLLIPV
20	orf84ng	LDKKVYDLYESA	EIVTNK	VKR	SKW	FYAL	PVILLIPV
		190	200	210	220	230	240
	orf84-1.pep	ESATEQQA	VLDPK	TEGEP	VNN	GNLT	ADMVPTLSEK
25	orf84ng	ESATEQQA	VLDPK	TEGES	VNN	GNLT	ADMVPTLSEK
		250	260	270	280	290	300
	orf84-1.pep	EGGRTGC	CYSHQ	G	TALKE	VT	ELMCKDYVKNGLPFN
30	orf84ng	EGGRTGC	CYSHQ	G	TALKE	VT	ELMCKDYVKNGLPFN
		310	320	330	340	350	360
	orf84-1.pep	ATLGGK	PKQNL	MYDN	WEERG	KPFEG	IGGGVVG
35	orf84ng	ATLGGK	PKQNL	MYDN	WEERG	KPFEG	IGGGVVG
		370	380	390			
40	orf84-1.pep	ATLGGK	PKQNL	MYDN	WEERG	KPFEG	IGGGVVG
	orf84ng	ATLGGK	PKQNL	MYDN	WEERG	KPFEG	IGGGVVG

Based on this analysis, including the presence of a putative transmembrane domain (single-underlined) in the gonococcal protein, and a putative ATP/GTP-binding site motif A (P-loop, double-underlined), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 39

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 327>:

	1	GTGGTTTTCC	TGAATGCCGA	CAACGGGATA	TTGGTTCAGG	ACTTGCCTTT
50	51	TGAAGTCAAA	CTGAAAAAAT	TCCATATCGA	TTTTTACAAT	ACGGGTATGC
	101	CGCGTGATTT	CGCCAGCGAT	ATTGAAGTGA	CGGACAAGGC	AACCGGTGAG
	151	AAACTCGAGC	GCACCATCCG	CGTGAACCAT	CCTTTGACCT	TGCACGGCAT
	201	CACGATTTAT	CAGGCGAGTT	TTGCCGACGG	CGGTTCCGAT	TTGACATTCA
55	251	AGGCGTGGAA	TTTGGGTGAT	GCTTCGCGCG	AGCCTGTCGT	GTGAAGGCA
	301	ACATCCATAC	ACCAGTTTCC	GTTGGAAT	GGCAAACACA	AATATCGTCT
	351	TGAGTTCGAT	CAGTTCACCT	CTATGAATGT	GGAGGACATG	AGCGAGGGCG
	401	CGGAACGGGA	AAAAAGCCTG	AAATCCACGC	TGCCCGATGT	CCGCGCCGTT
	451	ACTCAGGAAG	GTCACAAATA	CACCAAT...	.....	.....TACCG
	501	TATCCGTGAT	GCGCCAGGCC	AGGCGGTCGA	ATATAAAAAC	TATATGCTGC
60	551	CGGTTTTGCA	GGAACAGGAT	TATTTTTGGA	TTACCGGCAC	GCGCAGCGC.



5  
10

```

601 TTGCAGCAGC AATACCGCTG GCTGCGTATC CCCTTGGACA AGCAGTTGAA
651 AGCGGACACC TTTATGGCAT TGCCTGAGTT TTTGAAAGAT GGGGAAGGGC
701 GCAAACGTCT .GTTGCCGAC GCAACCAAAG GCGCACCTGC CGAAATCCGC
751 GAACAATTCA TGCTGGCTGC GGAAACACG CTGAACATCT TTGCACAAAA
801 AGGCTATTTG GGATTGGACG AATTTATTAC GTCCAATATC CCGAAAGAGC
851 AGCAGGATAA GATGCAGGGC TATTCTACG AAATGCTTTA CGGCGTGATG
901 AACGCTGCTT TGGATGAAAC CAT.ACCCGG TACGGCTTGC CCGAATGGCA
951 GCAGGATGAA GCGCGGAATC GTTCCTGCT GCACAGTATG GATGCGTACA
1001 CCGGTTTGAC CGAATATCCC GCGCCTATGC TGCTGCAACT TGATGGGTTT
1051 TCCGAGGTGC GTTCGTCGGG TTTGCAGATG ACCCGTTCCC C.GGTCCGCT
1101 TTTGGTCTAT CTC...
```

This corresponds to the amino acid sequence <SEQ ID 328; ORF88>:

15  
20

```

1 MVFLNADNGI LVQDLPEFEV LKKFHIDFYN TGMPRDFASD IEVTDKATGE
51 KLERTIRVNH PLTLHGITIY QASFADGSD LTFKAWNLDG ASREPVVLKA
101 TSIHQFPLEI GKHKYRLEFD QFTSMNVEDM SEGAEREKSL KSTLPDVRVAV
151 TQEGHKYTNX XXXXXYRIRD APGQAVEYKN YMLPVLQEQD YFWITGTRSX
201 LQQQYRWLRI PLDKQLKADT FMALREFLKD GEGRKRXVAD ATKGAPAEIR
251 EQFMLAAENT LNIFAQKGYL GLDEFITSNI PKEQDKMQG YFYEMLYGVM
301 NAALDETXTR YGLPEWQQDE ARNRFLHSM DAYTGLTEYP APMLQLDQDF
351 SEVRSSGLQM TRSXGPPLLVI L...
```

Further work revealed the complete nucleotide sequence <SEQ ID 329>:

25  
30  
35  
40  
45  
50  
55  
60

```

1 ATGAGTAAAT CCCGTAGATC TCCCCCACTT CTTTCCCGTC CGTGGTTCCGC
51 TTTTTTCAGC TCCATGCGCT TTGCAGTCGC TTTGCTCAGT CTGCTGGGTA
101 TTGCATCGGT TATCGGTACG GTGTTCAGC AAAACCAGCC GCAGACGGAT
151 TATTTGGTCA AATTCGGATC GTTTTGGGCG CAGATTTTTG GTTTTCTGGG
201 ACTGTATGAC GTCTATGCTT CGGCATGGTT TGTCGTTATC ATGATGTTTT
251 TGGTGGTTTC TACCACTTTG TGCTGATTG GCAATGTGCC GCCGTTCTGG
301 CGCGAAATGA AGTCTTTTCG GGAAAAGGTT AAAGAAAAAT CTCTGGCGGC
351 GATGCGCCAT TCTTCGCTGT TGGATGTAAG AATTGCGCCC GAGGTGGCCA
401 AACGTTATCT GGAAGTACAA GGTTTTCAGG GAAAAACCAT TAACCGTGAA
451 GACGGGTCGG TTCTGATTGC CGCCAAAAAA GGCACAATGA ACAAATGGGG
501 CTATATCTTT GCCCATGTTG CTTTGATTGT CATTTGCCTG GCGGGTTGA
551 TAGACAGTAA CCTGCTGTTG AAAGTGGGTA TGCTGACCGG TCGGATGTTT
601 CCGGACAATC AGGCGGTTTA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT
651 GGGTGCCTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC
701 AGAGTGCCTC TGTGGTTTTC CTGAATGCCG ACAACGGGAT ATGTTTCAG
751 GACTTGCTCT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTACAA
801 TACGGGTATG CCGCGTGATT TCGCCAGCGA TATTGAAGTG ACGGACAAAG
851 CAACCGGTGA GAACTCGAG CGCACCATCC GCGTGAACCA TCCTTTGACC
901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTTGCCGACG GCGGTTCCGA
951 TTTGACATTC AAGGCGTGGA ATTTGGGTGA TGCTTCGCGC GAGCCTGTCT
1001 TGTTGAAGGC AACATCCATA CACCACTTTC CGTTGGAAT TGGCAAACAC
1051 AAATATCGTC TTGAGTTCGA TCAGTTCACT TCTATGAATG TGGAGGACAT
1101 GAGCGAGGGC GCGGAACGGG AAAAAAGCCT GAAATCCACG CTGAACGATG
1151 TCCGCGCCGT TACTCAGGAA GGTAAGAAAT ACACCAATAT CGGCCCTTCC
1201 ATTGTTTACC GTATCCGTGA TGCGGCAGGG CAGGCGGTCT AATATAAAAA
1251 CTATATGCTG CCGGTTTTCG AGGAACAGGA TTATTTTGG ATTACCGGCA
1301 CGCGCAGCGG CTTCAGCAGC CAATACCGCT GGCTGCGTAT CCCCTTGGAC
1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAAGA
1401 TGGGGAAGGG CGCAACGTC TGTTTGCCGA CGCAACCAAA GGCGCACCTG
1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CGGAAAACAC GCTGAACATC
1501 TTTGCACAAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCCAATAT
1551 CCGGAAAGAG CAGCAGGATA AGATGCAGGG CTATTTCTAC GAAATGCTTT
1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACGGCTTG
1651 CCCGAATGGC AGCAGGATGA AGCGCGGAAT CGTTTCCTGC TGCACAGTAT
1701 GGATGCGTAC ACGGGTTTGA CCGAATATCC CGCGCCTATG CTGCTGCAAC
1751 TCGATGGGTT TTCCGAGGTG CGTTCGTCGG GTTTCAGAT GACCCGTTC
1801 CCGGGTGC GC TTTTGGTCTA TCTCGGCTCG GTGCTGTTGG TATTGGGTAC
1851 GGTATTGATG TTTTATGTGC GCGAAAAACG GGCGTGGGTA TTGTTTTTTCAG
1901 ACGGCAAAAT CCGTTTTCGC ATGTCTTCGG CCGCAGCGA ACGGGATTTG
1951 CAGAAGGAAT TTCCAAAACA CGTCGAGAGT CTGCAACGGC TCGGCAAGGA
2001 CTTGAATCAT GACTGA
```

This corresponds to the amino acid sequence <SEQ ID 330; ORF88-1>:

65

```

1 MSKSRSPPEL LSRPWFAPFS SMRFAVALLS LLGIASVIGT VLQONQPD
51 YLVKFGSFWA QIFGFLGLYD VYASAWFVVI MMFLVSTSL CLIRNVPPFW
```

-220-

101 REMKSFREKV KEKSLAAMRH SLLDVKIAP EVAKRYLEVQ GFQGTINRE  
 151 DGSVLIAAKK GTMNKGYIF AHVALIVICL GGLIDSNLLL KLGMLTGRIV  
 201 PDNQAVYAKD FKPEILGAS NLSFRGNVNI SEGQADVVF LNADNGILVQ  
 251 DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPLT  
 5 LHGITYIQAS FADGGSDLTF KAWNLDASR EPVVLKATSI HQFPLEIGKH  
 351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKYTNIGPS  
 401 IVYRIRDAAG QAVEYKNYML PVLQEQQDYFW ITGTRSGLQQ QYRWLRIPLD  
 451 KQLKADTFMA LREFLDGEG RKRLVADATK GAPAEIREQF MLAAENTLNI  
 501 FAQKGYLGLD EFITSNIPKE QQDKMQGYFY EMLYGMNAA LDETIRRYGL  
 10 PEWQQDEARN RFLHSMDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTRS  
 601 PGALLVYLGS VLLVLGTVLM FYVREKRAWV LFSDGKIRFA MSSARSERDL  
 651 QKEFPKHVES LQRLGKDLNH D\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

15 ORF88 shows 95.7% identity over a 371aa overlap with an ORF (ORF88a) from strain A of *N. meningitidis*:

	orf88.pep				10	20	30
					MVFLNADNGILVQDLPEFVKLKKFHIDFYN		
20	orf88a	AKDFKPESILGASNLSFRGNVNI	SEGQADVFLNADNGILVQDLPEFVKLKKFHIDFYN				
		210	220	230	240	250	260
	orf88.pep		40	50	60	70	80
25		TGMPRDFASDIEVTDKATGEKLE	RTIRVNHPLTLHGITYIQAS	FADGGSDLTFKAWNLDG			
	orf88a	TGMPRDFASDIEVTDKATGEKLE	RTIRVNHPLTLHGITYIQAS	FADGGSDLTFKAWNLDG			
		270	280	290	300	310	320
30	orf88.pep	ASREPVLKATSIHQFPLEIGKH	KYRLEFDQFTSMNVEDMSEGA	EEREKSLKSTLPDVR			
	orf88a	ASREPVLKATSIHQFPLEIGKH	KYRLEFDQFTSMNVEDMSEGA	EEREKSLKSTLNDVR			
		330	340	350	360	370	380
35	orf88.pep		160	170	180	190	200
		TQEGHKYTNXXXXXX	YRIRDAPQAVEYKNYML	PVLQEQQDYFWITG	TRSGLQQQYRWLR		
	orf88a	TQEGKKYTNIGPS	IYRIRDAAGAVEYKNYML	PVLQEQQDYFWITG	TRSGLQQQYRWLR		
		390	400	410	420	430	440
40	orf88.pep		220	230	240	250	260
		PLDKQLKADTFMALREFLDG	EGRKRVADATKGAPAEIREQ	FMLAAENTLNIFAQK	GYL		
	orf88a	PLDKQLKADTFMALREFLDG	EGRKRLVADATKGAPAEIREQ	FMLAAENTLNIFAQK	GYL		
45		450	460	470	480	490	500
50	orf88.pep		280	290	300	310	320
		GLDEFITSNIPKEQQDKMQGYFY	EMLYGMNAAALDET	XTRYGLPEWQQDEARN	RFLHSM		
	orf88a	GLDEFITSNIPKEQQDKMQGYFY	EMLYGMNAAALDET	IRRYGLPEWQQDEARN	RFLHSM		
		510	520	530	540	550	560
55	orf88.pep		340	350	360	370	
		DAYTGLTEYPAPMLLQLDGF	SEVRSSGLQMTRSXP	LLVYL			
	orf88a	DAYTGLTEYPAPMLLQLDGF	SEVRSSGLQMTRSPG	ALLVYLGSVLLVLGTVLMFYVREKR			
		570	580	590	600	610	620
60	orf88a	AWVLFSDGKIRFAMSSARSERDL	QKEFPKHVESLQRLGKDLNHDX				
		630	640	650	660	670	

The complete length ORF88a nucleotide sequence <SEQ ID 331> is:

1 ATGAGTAAAT CCCGTAGATC TCCCCACTT CTTTCCCGTC CGTGGTTCGC  
 51 TTTTTCAGC TCCATGCGCT TTGCGGTCGC TTGCTCAGT CTGCTGGGTA  
 101 TTGCATCGGT TATCGGTACG GTGTTGCAGC AAAACCAGCC GCAGACGGAT

-221-

151 TATTTGGTCA AATTCGGATC GTTTTGGGCG CAGATTTTGG GTTTTCTGGG  
 201 ACTGTATGAC GTCTATGCTT CGGCATGGTT TGTCGTTATC ATGATGTTTT  
 251 TGGTGGTTTC TACCAGTTTG TGCCTGATTC GCAATGTGCC GCCGTCTCTGG  
 301 CGCGAAATGA AGTCTTTTCG GGAAAAGGTT AAAGAAAAAT CTCTGGCGGC  
 351 GATGCGCCAT TCTTCGCTGT TGGATGTAAA AATTGCGCCC GAGGTTGCCA  
 401 AACGTTATCT GGAAGTACAA GGTTTTCAGG GAAAAACCAT TAACCGTGAA  
 451 GACGGGTCGG TTCTGATTGC CGCCAAAAAA GGCACAATGA ACAAATGGGG  
 501 CTATATCTTT GCCCATGTTG CTTTGATTGT CATTTCCTCG GCGCGGTTGA  
 551 TAGACAGTAA CCTGCTGTTG AAAC TGGGTA TGCTGACCGG TCGGATTGTT  
 601 CCGGACAATC AGGCGGTTTA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT  
 651 GGGTGCCTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC  
 701 AGAGTGCCTG TGTGGTTTTT CTGAATGCCG ACAACGGGAT ATTGGTTCAG  
 751 GACTTGCCCTT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTACAA  
 801 TACGGGTATG CCGCGCGATT TTGCCAGTGA TATTGAAGTA ACGGATAAGG  
 851 CAACCGGTGA GAACTCGAG CGCACCATCC GCGTGAACCA TCCTTTGACC  
 901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTTGCCGACG GCGGTTCCGA  
 951 TTTGACATTC AAGGCGTGGA ATTTGGGTGA TGCTTCGCGC GAGCCTGTCTG  
 1001 TGTTGAAGGC AACATCCATA CACCAGTTTC CGTTGGAAAT TGGCAAACAC  
 1051 AAATATCGTC TTGAGTTTCA TCAGTTTACT TCTATGAATG TGGAGGACAT  
 1101 GAGCGAGGGC GCGGAACGGG AAAAAAGCCT GAAATCCACG CTGAACGATG  
 1151 TCCGCGCCGT TACTCAGGAA GGTA AAAAAT ACACCAATAT CCGCCCTTCC  
 1201 ATTGTTTACC GTATCCGTGA TGCGGCAGGG CAGGCGGTCTG AATATAAAAA  
 1251 CTATATGCTG CCGGTTTTTG AGGAACAGGA TTATTTTGG ATTACCGGCA  
 1301 CGCGCAGCGG CTTGCAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGGAC  
 1351 AAGCAGTTGA AAGCGGCAC CTTTATGGCA TTGCGTGAGT TTTTGAAAGA  
 1401 TGGGGAAGGG CGCAAACGTC TGGTTGCCGA CGCAACCAA GGCGCACCTG  
 1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CGGAAAACAC GCTGAACATC  
 1501 TTTGCACAAA AAGGCTATTT GGGATTGAC GAATTTATTA CGTCCAATAT  
 1551 CCCGAAAGAG CAGCAGGATA AGATGCAGGG CTATTTCTAC GAAATGCTTT  
 1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACCGTTG  
 1651 CCCGAATGGC AGCAGGATGA AGCGCGGAAT CGTTTCCTGC TGCACAGTAT  
 1701 GGATGCGTAC ACGGGTTTGA CCGAATATCC CGCGCCTATG CTGCTGCAAC  
 1751 TTGATGGGTT TTCCGAGGTG CGTTCGTCGG GTTTGCAGAT GACCCGTTCC  
 1801 CCGGGTGCGC TTTTGGTCTA TCTCGGCTCG GTGCTGTTGG TATTGGGTAC  
 1851 GGTATTGATG TTTTATGTC GCGAAAACG GCGGTGGGTA TTGTTTTCAG  
 1901 ACGGCAAAAT CCGTTTTGCC ATGTCTTCGG CCCGCAGCGA ACGGGATTG  
 1951 CAGAAGGAAT TTCCAAAACA CGTCGAGAGT CTGCAACGGC TCGGCAAGGA  
 2001 CTTGAATCAT GACTGA

This encodes a protein having amino acid sequence <SEQ ID 332>:

40 1 MSKSRRSPPL LSRPWFAPFS SMRFAVALLS LLGIASVIGT VLQONQPQTD  
 51 YLVKFGSFWA QIFGFLGLYD VYASAWFVVI MMFLVSTSL CLIRNVPPFW  
 101 REMKSFREKV KEKSLAAMRH SLLDVKIAP EVAKRYLEVQ GFQGTINRE  
 151 DGSVLIAAKK GTMNKWGYIF AHVALIVICL GGLIDSNLLL KLGLMTGRIV  
 201 PDNQAVYAKD FKPEILGAS NLSFRGNVNI SEGQSADVVF LNADNGILVQ  
 45 251 DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPLT  
 301 LHGITYIQAS FADGGSDLTF KAWNLDASR EPVVLKATSI HQFPLEIGKH  
 351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKTYNIGPS  
 401 IVYRIRDAAG QAVEYKNYML PVLQEQDYFW ITGTRSGLQQ QYRWLRIFLD  
 451 KQLKADTFMA LREFLKDGEG RKRLVADATK GAPAEIREQF MLAENTLNI  
 50 501 FAQKGYLGLD EFITSNIPKE QQDKMQGYFY EMLYGVMAA LDETIRRYGL  
 551 PEWQQDEARN RFLHSMDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTRS  
 601 PGALLVYLG S VLLVLGTVLM FYVREKRAVW LFSDGKIRFA MSSARSERDL  
 651 QKEFPKHVES LQRLGKDLNH D\*

ORF88a and ORF88-1 100.0% identity in 671 aa overlap:

55 orf88a.pep MSKSRRSPPLLSRPWFAPFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA 60  
 orf88-1 MSKSRRSPPLLSRPWFAPFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA 60  
 60 orf88a.pep QIFGFLGLYDVYASAWFVIMMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120  
 orf88-1 QIFGFLGLYDVYASAWFVIMMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120  
 orf88a.pep SLLDVKIAPEVAKRYLEVQGFQGTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL 180  
 orf88-1 SLLDVKIAPEVAKRYLEVQGFQGTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL 180  
 65 orf88a.pep GGLIDSNLLLKLGLMTGRIVPDNQAVYAKDFKPEILGASNLSFRGNVNISEGQSADVVF 240

	orf88-1	GGGLIDSNLLKLKGLMTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSAADVVF	240
5	orf88a.pep	LNADNGILVQDLPFVEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLETRIRVNHPLT	300
	orf88-1	LNADNGILVQDLPFVEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLETRIRVNHPLT	300
	orf88a.pep	LHGITIYQASFADGGSDLTFAKAWNLDASREPVVLKATSIHQFPLEIGKHKYRLEFDQFT	360
10	orf88-1	LHGITIYQASFADGGSDLTFAKAWNLDASREPVVLKATSIHQFPLEIGKHKYRLEFDQFT	360
	orf88a.pep	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
15	orf88-1	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
	orf88a.pep	PVLQEQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
	orf88-1	PVLQEQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
20	orf88a.pep	GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA	540
	orf88-1	GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA	540
25	orf88a.pep	LDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS	600
	orf88-1	LDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS	600
	orf88a.pep	PGALLVYLGSVLLVLGTVLMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
30	orf88-1	PGALLVYLGSVLLVLGTVLMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
	orf88a.pep	LQRLGKDLNHD	672
	orf88-1	LQRLGKDLNHD	672

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF88 shows 93.8% identity over a 371aa overlap with a predicted ORF (ORF88.ng) from *N.*

*gonorrhoeae:*

40	orf88.pep	MVFLNADNGILVQDLPFVEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNH        :	60
	orf88ng	MVFLNADNGMLVQDLPFVEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNH	60
45	orf88.pep	PLTLHGITIYQASFADGGSDLTFKAWN LGDASREPVVLKATSIHQFPLEIGKHKYRLEFD        :	120
	orf88ng	PLTLHGITIYQASFADGGSDLTFKAWNLRDASREPVVLKATSIHQFPLEIGKHKYRLEFD	120
50	orf88.pep	QFTSMNVEDMSEGAEREKSLKSTLPDVRVQTQEGHKYTNXXXXXXYRIRDAPGQAVEYKN        :	180
	orf88ng	QFTSMNVEDMSEGAEREKSLKSTLNDVRVQTQEGKKYTNIGPSIVYRIRDAAGQAVEYKN	180
55	orf88.pep	YMLPVLQE QDYFWITGTRSLQQQYRWLRIP LDKQLKADTFMALREF LKDGEGRKRXVAD    : : : : :	240
	orf88ng	YMLPILQDKDYFWLTGTRSGLQQQYRWLRIP LDKQLKADTFMALREF LKDGEGRKRLVAD	240
60	orf88.pep	ATKGAPAEIREQFM LAAENTLNI FAQKG YLGLDEFITSNIPKEQQDKMQGYFYEMLYGVM 	300
	orf88ng	ATKAPAEIREQFM LAAENTLNI FAQKG YLGLDEFITSNIPKGQQDKMQGYFYEMLYGVM	300
65	orf88.pep	NAALDET XTRYGLPEWQQDEARNRFL LHSM DAYTGLTEY PAMLLQLDGFSEVRSSGLQM 	360
	orf88ng	NAALDETIRRYGLPEWQQDEARNRFL LHSM DAYTGLTEY PAMLLQLDGFSEVRSSGLQM	360
70	orf88.pep	TRSXGPLL VYL 	371
	orf88ng	TRSPGALLVYLGSVLLVLGTVFMFYVPKKRAWVLF SNXKIRFAMSSARSERDLQKEFPKH	420

An ORF88ng nucleotide sequence <SEQ ID 333> was predicted to encode a protein having amino acid sequence <SEQ ID 334>:

```

1  MVFLNADNGM LVQDLPEFEVK LKKFHIDFYN TGMPRDFASD IEVTDKATGE
5  51  KLERTIRVNH PLTLHGITYY QASFADGGSD LTFKAWNLRD ASREPVVLKA
101 TSIHQFPLEI GKHKYRLEFD QFTSMNVEDM SEGAEREKSL KSTLNDVRAV
151 TQEGKKYTNI GPSIVYRIRD AAGQAVEYKN YMLPILQDKD YFWLTGTRSG
201 LQQQYRWLRI PLDKQLKADT FMALREFLKD GEGRKRLVAD ATKDAPAEIR
251 EQFMLAAENT LNIFAQKGYL GLDEFITSNI PKGQQDKMQG YFYEMLYGVM
10 301 NAALDETIRR YGLPEWQQDE ARNRFLHSM DAYTGLTEYP APMLLQLDGF
351 SEVRSSGLQM TRSPGALLVY LGSVLLVLGT VFMFVYPKKR AWWLFSNXXKI
401 RFAMSSARSE RDLQKEFPKH VESLQRLGKD LNHD*

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 335>:

```

1  ATGAGTAAAT CCCGTATATC TCCCACACTT CTTTCCCGTC CGTGGTTTCGC
15 51  TTTTTCAGC TCCATGCGCT TTGCGGTCGC TTTGCTCAGT CTGCTGGGTA
101 TTGCATCGGT TATCGGCAGC GTGTTACAGC AAAACCAGCC GCAGACGGAT
151 TATTTGGTCA AATTCCGACC GTTTTGGACT CGGATTTTGG ATTTTGGGG
201 TTTGTATGAT GTCTATGCTT CGGCATGGTT TGTCGTTATC ATGATGTTTC
251 TGGTGGTTTC TACCAGTTTG TGTTTAATCC GTAACGTTCC GCCGTTTGG
20 301 CGCGAAATGA AGTCTTTCCG GGAAAAGGTT AAAGAAAAT CTCTGGCGGC
351 GATGCGCCAT TCTTCGCTGT TGGATGTAAA AATTGCCCCC GAAGTTGCCA
401 AACGTTATCT GGAGGTGCGG GGTTTTCAGG GAAAAACCGT CAGCCGTGAG
451 GACGGGTCGG TTCTGATTGC CGCCAAAAAA GGC Aaatga acaaATGGGG
501 CTATATCTTT GCccaagtag ctTTGATTGT CATTGCGCTG GCGGGTTGA
25 551 TAGACAGTAA CCTGCTGCTG AAGCTGGGTA TGCTGGCCGG TCGGATTGTT
601 CCGGACAATC AGCGGTTTGA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT
651 GGGTGCCTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC
701 AAAGTGCAGA TGTGGTTTTC CTGAATGCCG ACAACGGGAT GTTGGTTCAG
751 GACTTGCCCTT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTACAA
801 TACGGGTATG CCGCGCGATT TTGCCAGCGA TATTGAAGTA ACGGACAAGG
30 851 CAACCGGTGA GAAACTCGAG CGCACCATCC GCGTGAACCA TCCTTTGACC
901 TTGCACGGCA TCACGATTGA TCAGGCGAGT TTTGCCGACG GCGGTTCCGA
951 TTTGACATTC AAGGCGTGGA ATTTGAGGGA TGCTTCGCGC GAACCTGTCTG
1001 TGTTGAAGGC AACCTCCATA CACCAGTTTC CGTTGGAAAT CCGCAAAACAC
1051 AAATATCGTC TTGAGTTTCA TCAGTTCACT TCTATGAATG TGGAGGACAT
35 1101 GAGCGAGGGT GCGGAACGGG AAAAAAGCCT GAAATCCACT CTGAACGATG
1151 TCCGCGCCGT TACTCAGGAA GGTA AAAAAT ACACCAATAT CGGCCCTTCC
1201 ATCGTGTAAC GCATCCGTGA TGcggCAGGG CAGGCGGTCG AATATAAAAA
1251 CTATATGCTG CCGATTTTGC AGGACAAAGA TTATTTTGG CTGACCGGCA
40 1301 CGCGCAGCGG CTTGCAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGGAC
1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAGA
1401 TGGGAAGGG CGCAAACGTC TGGTTGCCGA CGCAACCAA GACGCACCTG
1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CGGAAAACAC GCTGAATATC
1501 TTTGCGCAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCCAATAT
45 1551 CCCGAAAGGG CAGCAGGATA AGATGCAGGG CTATTTCTAC GAAATGCTTT
1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACGGCTTG
1651 CCCGAATGGC AGCAGGATGA AGCGCGGAAC CGTTTCCTGC TGCACAGTAT
1701 GGATGCCTAT ACGGGGCTGA CGGAATATCC CGCGCCTATG CTGCTCCAGC
1751 TTGACGGGTT TTCCGAGGTG CGTTCCTCAG GTTTGCAGAT GACCCGTTTCG
50 1801 CCGGGTGCGC TTTTGGTCTA TCtgggctcg gtattgttg TTTTGGgtac
1851 ggtaTttatg tTTTATGTGC GCGAAAACG GCGGTGGgta tTGTTTTcag
1901 aCGGCAAAAT CCGTTTTGCT ATGtCTTcgg CCcgcagcga ACGGGATTTG
1951 cAGAaggaat TTCCAAAACA CGtcgAGAGC CTGCAACggc tcggcaaggA
2001 CttgaatCAT GACTga

```

This corresponds to the amino acid sequence <SEQ ID 336; ORF88ng-1>:

```

55 1  MSKSRIPTL LSRPWFAPFS SMRFAVALLS LLGIASVIGT VLQONQPQTD
51 51  YLVKFGPFWT RIFDFLGLYD VYASAWFVVI MMFLVSTSL CLIRNVPPFW
101 REMKSFREKV KEKSLAAMRH SSLLDVKIAP EVAKRYLEVR GFQKTVSRE
151 DGSVLIAAKK GTMNKWGYIF AQVALIVICL GGLIDSNLLL KLGMLAGRIV
60 201 PDNQAVYAKD FKPEILGAS NLSFRGNVNI SEGQSAADVVF LNADNGMLVQ
251 DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPLT
301 LHGITYYQAS FADGGSDLTF KAWNLRDASR EPVVLKATSI HQFPLEIGKH
351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKYTNIGPS
401 IVYRIRDAAG QAVEYKKNYML PILQDKDYFW LTGTRSGLQQ QYRWLRIPLD
451 KQLKADTFMA LREFLKDGEF RKRLVADATK DAPAEIREQF MLAENTLNI
65 501 FAQKGYLGLD EFITSNIPKG QQDKMQGYFY EMLYGMNAA LDETIRRYGL

```

551 PEWQQDEARN RFLHLSMDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTRS  
601 PGALLVYLGS VLLVLGTVFM FYVREKRAWV LFS DGKIRFA MSSARSERDL  
651 OKEFPKHVES LORLGKDLNH D\*

ORF88ng-1 and ORF88-1 show 97.0% identity in 671 aa overlap:

5	orf88-1.pep	MSKSRRSPLLRSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA	60
	orf88ng-1	MSKSRISPTLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGFPWT	60
10	orf88-1.pep	QIFGFLGLYDVYASAWFVIMMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	120
	orf88ng-1	RIFDFLGLYDVYASAWFVIMMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	120
15	orf88-1.pep	SLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKGWYIFAHVALLIVICL	180
	orf88ng-1	SLLDVKIAPEVAKRYLEVGRGFQGKTVSREDGSVLIAAKKGTMNKGWYIFAQVALLIVICL	180
20	orf88-1.pep	GGLIDSNLLKLGLMLTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF	240
	orf88ng-1	GGLIDSNLLKLGLMLAGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF	240
25	orf88-1.pep	LNADNGILVQDLPFEVCLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLETRIRNVHPLT	300
	orf88ng-1	LNADNGMLVQDLPFEVCLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLETRIRNVHPLT	300
30	orf88-1.pep	LHGITIYQASFADGGSDLTFKAWNLDASREPVLKATSIHQFPLEIGKHKYRLEFDQFT	360
	orf88ng-1	LHGITIYQASFADGGSDLTFKAWNLRDASREPVLKATSIHQFPLEIGKHKYRLEFDQFT	360
35	orf88-1.pep	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
	orf88ng-1	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
40	orf88-1.pep	PVLQEQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
	orf88ng-1	PILQDKDYFWLTGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
45	orf88-1.pep	GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA	540
	orf88ng-1	DAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKGQQDKMQGYFYEMLYGVMNAA	540
50	orf88-1.pep	LDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS	600
	orf88ng-1	LDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS	600
55	orf88-1.pep	PGALLVYLGSVLLVLGTVLMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
	orf88ng-1	PGALLVYLGSVLLVLGTVFMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
60	orf88-1.pep	LQRLGKDLNHD 671	
	orf88ng-1	LQRLGKDLNHD 671	

Furthermore, ORG88ng-1 shows homology with a hypothetical protein from *Aquifex aeolicus*:

```

gi|2984296 (AE000771) hypothetical protein [Aquifex aeolicus] Length = 537
Score = 94.4 bits (231), Expect = 2e-18
Identities = 91/334 (27%), Positives = 159/334 (47%), Gaps = 59/334 (17%)

55 Query: 16  FAFSSMRFAVALLSLLGIASVIG-TVLQONQPQT DYLVKFGPFWTRIFDFLGLYDVYAS 74
      + F +S++ A+ ++ +LGI S++G T ++QNQ      YL +FG      L L DV+ S
Sbjct: 80  YDFLASLKLAI FIMLVLGILSMLGSTYIKQNQSFEWYLDQFGYDVGVIWKLWLN DNVFHS 139

60 Query: 75  AWFVVIMMFLV VSTSLCLIRNVPPFWREMKS FREKVKEKSLAAMRHSSLLDV KIAPEVAK 134
      +++++ ++ L V+   C I+ +P W++ S +E++ +   A +H   + VKI P+ K
Sbjct: 140 WYYILFIVLLAVNLIFCSIKRLPRVWQQA FS-KERILKLDEHAEKHLKPITVKI-PDKDK 197

65 Query: 135 --RYLEVRGFGQKTVSREDGSVLIAAKKGT MNKWGYIFAQVALIVICLGLIDS NLLKL 192
      ++L +GF+  V E  ++ A+KG ++ G      +AL+VI G LID
Sbjct: 198 VLKFL LKKGFK-VFVEEGNKLYVFAEKG RFSRLGVYITHIALLVIMAGALID----- 249

```

```

Query: 193 GMLAGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVFLNADNGMLVQDL 252
              +I+G      RG++ ++EG + DV+ + A+      L
Sbjct: 250 -----AIVGV-----RGSLLVAEGDTNDVMLVGAE--QKPYKL 280

Query: 253 PFEVKLKKFHIDFY---NTGMPRDFA-----SDIEVTDKATGEKLER--TIRVNHPLT 300
              PF V L F I Y N + + FA      SDIE+ + G K+E T++VN P
Sbjct: 281 PFAVHLIDFRIKTYAEENPNVDKRFQAQAVSSYESDIEIIN---GGKVEAKGTVKVNPF 337

Query: 301 LHGITIYQASFA--DGGSDLTFKAWNLRDASREP 332
              ++QA++ DG S + + + A +P
Sbjct: 338 FGRYRLFQATYGILDGTSGMGVIVVDRKKAHEDP 371

```

Based on this analysis, including the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 40

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 337>:

```

1  ATGATGAGTA ATAmAATGGm ACAAAAAGGG TTTACATTGA TTGmGmTGAT
51  GATAGTCGTC GCGATACTCG GCATTATCAG CGTCATTGCC ATACCTTCTT
101 ATCmAAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
151 GycGGTATCA ACAATATTTT CAAACAGTTT ATTTTGAAAA ATCCCCTGGA
201 CGATAATCAG ACCATCGAGA ACAAACCTGGA AATATTTGTC TCAGGCTATA
251 AGATGAATCC GAAAATTGCC AAAAAaTATA GTGTTTCGGT AAAGTTTGTC
301 GATAAGGAAA AATCAAGGGC ATACAGGTTG GTCGGCGTTC CGAAGGCGGG
351 GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA
401 AATGCCGTGA TGCCGCTTCT GCCCAAGCCC ATTTGGAGAC CTTGTCCTCA
451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAA

```

This corresponds to the amino acid sequence <SEQ ID 338; ORF89>:

```

1  MMSNXMXQKG FTLIXXMIVV AILGIISVIA IPSYXSYIEK GYQSQLYTEM
51  XGINNISKQF ILKNPLDDNQ TIENKLEIFV SGYKMNP KIA KKYSVSVK FV
101 DKEKSRAYRL VGVPKAGTGY TLSVWMNSVG DGYKCRDAAS AQAHLET LSS
151 DVGCEAFSNR KK*

```

Further work revealed the complete nucleotide sequence <SEQ ID 339>:

```

1  ATGATGAGTA ATAAAATGGA ACAAAAAGGG TTTACATTGA TTGAGATGAT
51  GATAGTCGTC GCGATACTCG GCATTATCAG CGTCATTGCC ATACCTTCTT
101 ATCAAAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
151 GTCGGTATCA ACAATATTTT CAAACAGTTT ATTTTGAAAA ATCCCCTGGA
201 CGATAATCAG ACCATCGAGA ACAAACCTGGA AATATTTGTC TCAGGCTATA
251 AGATGAATCC GAAAATTGCC AAAAAATATA GTGTTTCGGT AAAGTTTGTC
301 GATAAGGAAA AATCAAGGGC ATACAGGTTG GTCGGCGTTC CGAAGGCGGG
351 GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA
401 AATGCCGTGA TGCCGCTTCT GCCCAAGCCC ATTTGGAGAC CTTGTCCTCA
451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAA

```

This corresponds to the amino acid sequence <SEQ ID 340; ORF89-1>:

```

1  MMSNKM EQKG FTLIEMMIVV AILGIISVIA IPSYQSYIEK GYQSQLYTEM
51  VGINNISKQF ILKNPLDDNQ TIENKLEIFV SGYKMNP KIA KKYSVSVK FV
101 DKEKSRAYRL VGVPKAGTGY TLSVWMNSVG DGYKCRDAAS AQAHLET LSS
151 DVGCEAFSNR KK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with Pile of *N. gonorrhoeae* (accession number Z69260).

ORF89 and Pile protein show 30% aa identity in 120a overlap:

-226-

orf89 8 QKGFTLIXXMIVVAILGIISVIAIPSYXSYIEKGYQSOLYTEMXGINNISKQFILKNPL- 66  
 QKGFTLI MIV+AI+GI++ +A+P+Y Y + S+ G + ++ L + +  
 Pile 5 QKGFTLIELMIVIAIVGILAAVALPAYQDYTARAQVSEAILLAEGQKSAVTEYYLNHGIW 64

5 orf89 67 -DDNQTIENTKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAVRLVGVPKAGTGYTLVSW 125  
 DN + +G + KI KY SV + GV K G LS+W  
 Pile 65 PKDNTS-----AGVASSDKIKGKYVQSVTVAKGVVTAEMASTGVNKEIQGKKLSLW 115

### Homology with a predicted ORF from *N.meningitidis* (strain A)

10 ORF89 shows 83.3% identity over a 162aa overlap with an ORF (ORF89a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf89.pep		MMSNXMXQKGFTLIXXMIVVAILGIISVIAIPSYXSYIEKGYQSOLYTEMXGINNISKQF					
15 orf89a		MMSNKMEQKGFTLIXXXXXAIXXXXSVIXXXXSYIEKGYQSOLYTEMVGINNISKQX					
		10	20	30	40	50	60
		70	80	90	100	110	120
20 orf89.pep		ILKNPLDDNQTIENTKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAVRLVGVPKAGTGY					
orf89a		ILKNPLDDNQTIKSKLEIFVSGYKMNPKIAEKYNVSVHFVNEEKPRAYSLVGVPKTGTGY					
		70	80	90	100	110	120
		130	140	150	160		
25 orf89.pep		TLVWVMSVGDGYKCRDAASAQAHALETLSDDVGCEAFSNRKKX					
orf89a		TLVWVMSVGDGYKCRDAASARAHLETLSDDVGCEAFSNRKKX					
		130	140	150	160		

The complete length ORF89a nucleotide sequence <SEQ ID 341> is:

30 1 ATGATGAGTA ATAAAATGGA ACAAAAAGGG TTTACATTGA TTGNGANGNT  
 51 NATNGNCNTC GCGATACNCN GCNTTANCAG CGTCATTNCN ATNNNTNCNT  
 101 ATCENNAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG  
 151 GTCGGTATCA ACAATATTTC CAAACAGTNT ATTTTGAAAA ATCCCTTGGA  
 35 201 CGATAATCAG ACCATCAAGA GCAAACCTGGA AATATTGTC TCAGGCTATA  
 251 AGATGAATCC GAAAATTGCC GAAAAATATA ATGTTTCGGT GCATTTTGTC  
 301 AATGAGGAAA AACCNAGGGC ATACAGCTTG GTCGGCGTTC CAAAGACGGG  
 351 GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA  
 401 AATGCCGTGA TGCCGCTTCT GCCCGAGCCC ATTTGGAGAC CTTGTCTCTCA  
 451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAG

40 This encodes a protein having amino acid sequence <SEQ ID 342>:

1 MMSNKMEQKG FTLIXXXXXX AIXXXXSVIX XXXYXSYIEK GYQSOLYTEM  
 51 VGINNISKQX ILKNPLDDNQ TIKSKLEIFV SGYKMNPKIA EKYNVSVHFV  
 101 NEEKPRAYSL VGVPKTGTGY TSVWVMSVGD DGYKCRDAAS ARAHALETLS  
 151 DVGCEAFSNR KK\*

45 ORF89a and ORF89-1 show 83.3% identity in 162 aa overlap:

		10	20	30	40	50	60
orf89a.pep		MMSNKMEQKGFTLIXXXXXAIXXXXSVIXXXXSYIEKGYQSOLYTEMVGINNISKQX					
50 orf89-1		MMSNKMEQKGFTLIEMMIVVAILGIISVIAIPSYQSYIEKGYQSOLYTEMVGINNISKQF					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf89a.pep		ILKNPLDDNQTIKSKLEIFVSGYKMNPKIAEKYNVSVHFVNEEKPRAYSLVGVPKTGTGY					
55 orf89-1		ILKNPLDDNQTIENTKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAVRLVGVPKAGTGY					
		70	80	90	100	110	120
		130	140	150	160		
60 orf89a.pep		TLVWVMSVGDGYKCRDAASARAHLETLSDDVGCEAFSNRKKX					
orf89-1		TLVWVMSVGDGYKCRDAASAQAHALETLSDDVGCEAFSNRKKX					



-227-

130 140 150 160

Homology with a predicted ORF from *N.gonorrhoeae*ORF89 shows 84.6% identity over a 162aa overlap with a predicted ORF (ORF89.ng) from *N.*5 *gonorrhoeae*:

```

orf89      MMSNKMxQKGFTLIXMIVVAILGIISVIAIPSYXSIEKGYQSQLYTEMXGINNISKQF 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf89ng     MMSNKMEQKGFTLIEMMIVVTILGIISVIAIPSYQSYIEKGYQSQLYTEMVGINNVLKQF 60

10 orf89      ILKNPLDDNQTIENTKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGY 120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf89ng     ILKNPQDDNDTLKSKLKIFVSGYKMNPKIAKKYSVSVRFVDAEKPRAYRLVGVPNAGTGY 120

15 orf89      TLSVWMNSVGDGYKCRDAASAQAHALETLSSDVGCEAFSNRKK 162
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf89ng     TLSVWMNSVGDGYKCRDATSAQAYSDTLSADSGCEAFSNRKK 162

```

The complete length ORF89ng nucleotide sequence &lt;SEQ ID 343&gt; is:

```

1  aTGATGAGCA ATAAAATGGA ACAAAAAGGG TTTACATTGA TTGAGATGAT
20 51  GATAGTTGTC ACGATACTCG GCATCATCAG CGTCATTGCC ATACCTTCTT
101 ATCAGAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
151 GTCGGTATCA ACAATGTTCT CAAACAGTTT ATTTTGAAAA ATCCCCAGGA
201 CGATAATGAT ACCCTCAAGA GCAAACTGAA AATATTGTGC TCAGGCTATA
251 AGATGAATCC GAAAAttGCC AAAAAATATA GTGTTTCGGt aagggttGTC
25 301  gatGCGGAAA AACCAAGGGC ATACAGGTTG GTCGGCGTTC CGAACGCGGG
351 GACGGGTAT ACTTTGTCCG TATGGATGAA CAGCGTGGC GACGGATACA
401 AATGCCGTGA TGCCACTTCT GCCCAGGCCT ATTCGACAC CTTGTCCGCA
451 GATAGCGGCT GTGAAGCTTT CTCTAATCGT AAAAAATAG

```

This encodes a protein having amino acid sequence &lt;SEQ ID 344&gt;:

```

1  MMSNKMEQKG FTLIEMMIVV TILGIISVIA IPSYQSYIEK GYQSQLYTEM
30 51  VGINNVLKQF ILKNPQDDND TLKSKLKIFV SGYKMNPKIA KKYSVSVRFV
101 DAEKPRAYRL VGVPNAGTGY TLSVWMNSVG DGYKCRDATS AQAYSDTLSA
151 DSGCEAFSNR KK*

```

This gonococcal protein has a putative leader peptide (underlined) and N-terminal methylation site (NMePhe or type-4 pili, double-underlined). In addition, ORF89ng and ORF89-1 show 88.3%

35 identity in 162 aa overlap:

```

10 20 30 40 50 60
orf89-1.pep MMSNKMEQKGFTLIEMMIVVAILGIISVIAIPSYQSYIEKGYQSQLYTEMVGINNISKQF
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
40 orf89ng     MMSNKMEQKGFTLIEMMIVVTILGIISVIAIPSYQSYIEKGYQSQLYTEMVGINNVLKQF
10 20 30 40 50 60

70 80 90 100 110 120
orf89-1.pep ILKNPLDDNQTIENTKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGY
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45 orf89ng     ILKNPQDDNDTLKSKLKIFVSGYKMNPKIAKKYSVSVRFVDAEKPRAYRLVGVPNAGTGY
70 80 90 100 110 120

130 140 150 160
orf89-1.pep TLSVWMNSVGDGYKCRDAASAQAHALETLSSDVGCEAFSNRKKX
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
50 orf89ng     TLSVWMNSVGDGYKCRDATSAQAYSDTLSADSGCEAFSNRKKX
130 140 150 160

```

Based on this analysis, including the gonococcal motifs and the homology with the known Pile protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their

55 epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF89-1 (13.6kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 11A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera gave a positive result in the ELISA test., confirming that

5 ORF89-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 41

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 345>:

```

10      1  ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
      51  CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAGCCAA ATCCGTCAAA
     101  ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
     151  CGCCAAAAAG CCGAAGCCTA TGCGATTCCC TATTTCGATT TCCAACGTAT
     201  GACCGCATTG GCGGTCGGCA ACCCTTGGsG CACCG.GTCC GACG.GCAAA
     251  AACAAAGCGTT GGCCn.AGAA TTTCACCCC...
```

This corresponds to the amino acid sequence <SEQ ID 346; ORF91>:

```

15      1  MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRQNATQVLS ILKNGDANTA
     51  RQKAEAYAIP YDFQRM TAL AVGNPWXTXS DXQKQALAXE FQP...
```

Further work revealed the complete nucleotide sequence <SEQ ID 347>:

```

20      1  ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
     51  CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAGCCAA ATCCGTCAAA
    101  ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
    151  CGCCAAAAAG CCGAAGCCTA TGCGATTCCC TATTTCGATT TCCAACGTAT
    201  GACCGCATTG GCGGTCGGCA ACCCTTGGCG CACCGCGTCC GACGCGCAAA
    251  AACAAAGCGTT GGCCAAAGAA TTTCACACCC TGCTGATCCG CACCTATTCC
    301  GGCACGATGC TGAAATTAAA AAACGCCAAC GTCAACGTCA AAGACAATCC
    351  CATCGTCAAT AAAGGCGGCA AAGAAATCAT CGTCCGCGCC GAAGTCGGCG
    401  TACCCGGGCA AAAACCCGTC AACATGGACT TCACCACCTA CCAAAGCGGC
    451  GGTAATATACC GTACCTACAA CGTCGCCATC GAAGGCGCGA GCCTGGTTAC
    501  CGTGTACCGC AACCAATTCG GCGAAATTAT CAAAGCGAAA GGCGTGGACG
    551  GACTGATTGC CGAGTTGAAA GCCAAAAACG GCGGCAAATA A
```

30 This corresponds to the amino acid sequence <SEQ ID 348; ORF91-1>:

```

      1  MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRQNATQVLS ILKNGDANTA
     51  RQKAEAYAIP YDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
    101  GTMLKLKLAN VNVKDNPIVN KGGKEIIVRA EVGVPQKPV NMDFTTYQSG
    151  GKYRTYNVAI EGASLVTYR NQFGEIILKAK GVDGLIAELK AKNGGK*
```

35 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF91 shows 92.4% identity over a 92aa overlap with an ORF (ORF91a) from strain A of *N. meningitidis*:

```

40      orf91.pep      10      20      30      40      50      60
      MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
      orf91a          10      20      30      40      50      60
      MKKSSFISALGIGILSIGMAFAAPADAVNQIRQNATQVLSILKSGDANTARQKAEAYAIP

45      orf91.pep      70      80      90
      YDFQRM TALAVGNPWXTXS DXQKQALAXE FQP
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
      orf91a          70      80      90      100      110      120
      YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYS GTMLKLKLANVNVKDNPIVN
```

orf91a KGGKEIIVRAEVGVPQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYVRNQFGEI IKAK  
130 140 150 160 170 180

The complete length ORF91a nucleotide sequence <SEQ ID 349> is:

```

5      1 ATGAAAAAAT CCTCCTTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
      51 CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAACCAA ATCCGTCAAA
     101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA GCGGTGATGC CAACACCGCC
     151 CGCCAAAAAG CCGAAGCCTA TGCGATTCCC TATTTGATT TCCAACGTAT
    201 GACCGCATTG GCGGTCGGCA ACCCTTGGCG CACCGCGTCC GACGCGCAAA
    251 AACAGCGTGT GGCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTATTCC
    301 GGCACGATGC TGAATTATAA AAACGCCAAC GTCAACGTCA AAGACAATCC
    351 CATCGTCAAT AAAGCGGCA AAGAAATCAT CGTCCGCGCC GAAGTCGGCG
    401 TACCCGGGCA AAAACCCGTC AACATGGACT TCACCACCTA CCAAAGCGGC
    451 GGTAATATACC GTACCTACAA CGTCGCCATC GAAGGCGCGA GCCTGGTTAC
    501 CGTGTACCGC AACCAATTCT GCGAAATTAT CAAAGCGAAA GGCGTGGACG
    551 GACTGATTGC CGAGTTGAAG GCTAAAAACG GCAGCAAGTA A
  
```

This encodes a protein having amino acid sequence <SEQ ID 350>:

```

      1 MKKSSFISAL GIGILSIGMA FAAPADAVNQ IRQATQVLS ILKSGDANTA
     51 RQKAEAYAIP YFDFQRM TAL AVGNPWR TAS DAQKQALAKE FQTL LIR TYS
    101 GTMLKLKNAN VNVKDNPIVN KGGKEIIVRA EVGVPQKPV NMDFTTYQSG
    151 GKYRTYNVAI EGASLVTYR NQFGEI IKAK GVDGLIAELK AKNGSK*
  
```

ORF91a and ORF91-1 show 98.0% identity in 196 aa overlap:

```

      10      20      30      40      50      60
    orf91a.pep MKKSSFISALGIGILSIGMAFAAPADAVNQIRQATQVLSILKSGDANTARQKAEAYAIP
    orf91-1     MKKSSLISALGIGILSIGMAFAAPADAVSQIRQATQVLSILKNGDANTARQKAEAYAIP
      10      20      30      40      50      60
    orf91a.pep YFDFQRM TAL AVGNPWR TAS DAQKQALAKEFQTL LIR TYS GTMLKLKNANVNVKDNPIVN
    orf91-1     YFDFQRM TAL AVGNPWR TAS DAQKQALAKEFQTL LIR TYS GTMLKLKNANVNVKDNPIVN
      70      80      90     100     110     120
    orf91a.pep YFDFQRM TAL AVGNPWR TAS DAQKQALAKEFQTL LIR TYS GTMLKLKNANVNVKDNPIVN
    orf91-1     YFDFQRM TAL AVGNPWR TAS DAQKQALAKEFQTL LIR TYS GTMLKLKNANVNVKDNPIVN
      70      80      90     100     110     120
    orf91a.pep KGGKEIIVRAEVGVPQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYVRNQFGEI IKAK
    orf91-1     KGGKEIIVRAEVGVPQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYVRNQFGEI IKAK
      130     140     150     160     170     180
    orf91a.pep KGGKEIIVRAEVGVPQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYVRNQFGEI IKAK
    orf91-1     KGGKEIIVRAEVGVPQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYVRNQFGEI IKAK
      130     140     150     160     170     180
    orf91a.pep GVDGLIAELKAKNGSKX
    orf91-1     GVDGLIAELKAKNGGKX
      190
    orf91a.pep GVDGLIAELKAKNGSKX
    orf91-1     GVDGLIAELKAKNGGKX
      190
  
```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF91 shows 84.8% identity over a 92aa overlap with a predicted ORF (ORF91.ng) from *N. gonorrhoeae*:

```

    50      orf91.pep MKKSSLISALGIGILSIGMAFAAPADAVSQIRQATQVLSILKNGDANTARQKAEAYAIP 60
           :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
           orf91ng  VKKSSFISALGIGILSIGMAFASPADAVGQIRQATQVLTILKSGDAASARPKAEAYAVP 60

    55      orf91.pep YFDFQRM TAL AVGNPWXTXSDXQKQALAXEFQP 93
           |||||:|||||:|||||:|||||:|||||:|||||:|||||:
           orf91ng  YFDFQRM TAL AVGNPWR TAS DAQKQALAKEFQTL LIR TYS GTMLKFKNATVNVKDNPIVN 120
  
```

The complete length ORF91ng nucleotide sequence <SEQ ID 351> is predicted to encode a protein having amino acid sequence <SEQ ID 352>:

-230-

```

1 VKKSSFISAL GIGILSIGMA FASPADAVGQ IRQNATQVLT ILKSGDAASA
51 RPKAEAYAVP YFDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
101 GTMLKFKNAT VNVKDNPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG
151 GKYRTYNVAI EGTSLVTVYR NQFGEI I KAK GIDGLIAELK AKNGGK*

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 353>:

```

1 ATGAAAAAAT CCTCCTTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
51 CGGCATGGCA TTTGCCTCCC CGGCCGACGC AGTGGGACAA ATCCGCCAAA
101 ACGCCACACA GGTTTTGACC ATCTCAAAA GCGGCGACGC GGCTTCTGCA
10 GCGCCAAAAG CCGAAGCCTA TGC GTTCCC TATTTCGATT TCCAACGTAT
201 GACCGCATTG GCGGTCGGCA ACCCTTGGCG TACCGCGTCC GACGCGCAA
251 AACAAGCGTT GGCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTATTCC
301 GGCACGATGC TGAAATTCAA AAACGCGACC GTCAACGTCA AAGACAATCC
351 CATCGTCAAT AAGGGCGGCA AGGAAATCGT CGTCCGTGCC GAAGTCGGCA
401 TCCCCGGTCA GAAGCCCGTC AATATGGACT TTACCACCTA CCAAAGCGGC
15 GGCAAATACC GTACCTACAA CGTCGCCATC GAAGGCACGA GCCTGGTTAC
501 CGTGTACCGC AACCAATTCG GCGAAATCAT CAAAGCCAAA GGCATCGACG
551 GGCTGATTGC CGAGTTGAAA GCCAAAACG GCGGCAAATA A

```

This corresponds to the amino acid sequence <SEQ ID 354; ORF91ng-1>:

```

1 MKKSSFISAL GIGILSIGMA FASPADAVGQ IRQNATQVLT ILKSGDAASA
20 RPKAEAYAVP YFDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
101 GTMLKFKNAT VNVKDNPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG
151 GKYRTYNVAI EGTSLVTVYR NQFGEI I KAK GIDGLIAELK AKNGGK*

```

ORF91ng-1 and ORF91-1 show 92.3% identity in 196 aa overlap:

```

25 orf91-1.pep      10      20      30      40      50      60
      MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf91ng-1      MKKSSFISALGIGILSIGMAFASPADAVGQIRQNATQVLTILKSGDAASARPKAEAYAVP
      10      20      30      40      50      60

30 orf91-1.pep      70      80      90      100     110     120
      YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKLKNANVNVKDNPIVN
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf91ng-1      YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKFKNATVNVKDNPIVN
      70      80      90      100     110     120

35 orf91-1.pep      130     140     150     160     170     180
      KGGKEIIVRAEVGPVGPQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTVYRNQFGEI I KAK
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf91ng-1      KGGKEIIVRAEVGIPGQKPVNMDFTTYQSGGKYRTYNVAIEGTSLVTVYRNQFGEI I KAK
      130     140     150     160     170     180

40 orf91-1.pep      190
      GVDGLIAELKAKNGGKX
      |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
45 orf91ng-1      GIDGLIAELKAKNGGKX
      190

```

In addition, ORF91ng-1 shows homology to a hypothetical *E.coli* protein:

```

50 sp|P45390|YRBC ECOLI HYPOTHETICAL 24.0 KD PROTEIN IN MURA-RPON INTERGENIC
      REGION PRECURSOR (F211) >gi|606130 (U18997) ORF_f211 [Escherichia coli]
      >gi|1789583 (AE000399) hypothetical 24.0 kD protein in murZ-rpoN intergenic
      region [Escherichia coli]Length = 211

      Score = 70.6 bits (170), Expect = 6e-12
      Identities = 42/137 (30%), Positives = 76/137 (54%), Gaps = 6/137 (4%)

55 Query: 59 VPYFDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKFKNATVNVKDNPI 118
      +PY + AL +G +++A+ AQ++A F+ L + Y + + T + P
      Sbjct: 65 LPYVQVKYAGALVLGQYYKSATPAQREAYFAAFREYLKQAYGQALAMYHGQTYQIA--PE 122

60 Query: 119 VNKGGKEIV-VRAEVGIP-GQKPVNMDFTTYQSG--GKYRTYNVAIEGTSLVTVYRNQFG 174
      G K IV +R + P G+ PV +DF ++ G ++ Y++ EG S++T +N++G
      Sbjct: 123 QPLGDKTIVPIRVTIIDPNRPPVRLDFQWRKNSQTGNWQAYDMIAEGVSMITTKQNEWG 182

```

-231-

Query: 175 EIIKAKGIDGLIAELKA 191  
 +++ KGIDGL A+LK+  
 Sbjct: 183 TLLRTKGIDGLTAQLKS 199

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 42**

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 355>:

```

10      1  ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTCAAC
      51  CGCTTCGGCA CATCCTGCCA GCGAACCGTC CACTCAAAAC GAAACCGCTA
     101  TGATCACGCA TACCCTCATC TCAAAATACA GTTTTGnnnn nnnnnnnnnn
     151  nnnnnnnnnn nnGCCATAAA AAGCAAAGGG ATGGACATT TGGCCGTCAT
     201  CGACCATCAG GAAGCCGCAC GCCGAAACGG CTTAACGATG CAGCCGGCAA
     251  AAGTCATCGT CTTCCGGCAG CCCAAAGCCG GCACGCCGCT GATGGTCAAA
    15      301  GACCCCGCCT TCGCCCTGCA ACTGCCCTA CGCGTCCTCG TTACCGAAAC
     351  GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG
     401  GCAGCCGCAT CGGTTTCGAC GAAGTGGCAA ACACCTTGGC AAACGCCGAA
     451  AACTGATAC AAAAAACCGT AGGCGAATAA
  
```

This corresponds to the amino acid sequence <SEQ ID 356; ORF97>:

```

20      1  MKHILPLIAA SALCISTASA HPASEPSTQN ETAMITHTLI SKYSFGXXXX
      51  XXXXAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
     101  DPAFALQLPL RVLVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE
     151  KLIQKTVGE*
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 357>:

```

25      1  ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTCAAC
      51  CGCTTCGGCA CATCCTGCCA GCGAACCGTC CACCCAAAAC GAAACCGCTA
     101  TGACCACGCA TACCCTCACC TCAAAATACA GTTTTGACGA AACCGTCAGC
     151  CGCCTTGAAA CCGCCATAAA AAGCAAAGGG ATGGACATT TGGCCGTCAT
     201  CGACCATCAG GAAGCCGCCC GCCGAAACGG CTTAACGATG CAGCCGGCAA
    30      251  AAGTCATCGT CTTCCGGCAG CCCAAAGCCG GCACGCCGCT GATGGTCAAA
     301  GACCCCGCCT TCGCCCTGCA ACTGCCCTA CGCGTCCTCG TTACCGAAAC
     351  GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG
     401  GCAGCCGCAT CGGTTTCGAC GAAGTGGCAA ACACCTTGGC AAACGCCGAA
     451  AACTGATAC AAAAAACCGT AGGCGAATAA
  
```

35 This corresponds to the amino acid sequence <SEQ ID 358; ORF97-1>:

```

      1  MKHILPLIAA SALCISTASA HPASEPSTQN ETAMTHTLT SKYSFDETVS
      51  RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
     101  DPAFALQLPL RVLVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE
     151  KLIQKTVGE*
  
```

40 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF97 shows 88.7% identity over a 159aa overlap with an ORF (ORF97a) from strain A of *N.meningitidis*:

```

45      orf97.pep      10      20      30      40      50      60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      orf97a      MXHILPLXXASALCISTASXHPASEPQTNETAMTHTLT SKYSFDETVSRLETAIKSKG
                  10      20      30      40      50      60
  
```

-232-

		70	80	90	100	110	120
	orf97.pep	MDIFAVIDHQAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVTETDGK					
5	orf97a	MDIFAVIDHQAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVXVTETDGK					
		70	80	90	100	110	120
	orf97.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					
10	orf97a	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					
		130	140	150	160		

The complete length ORF97a nucleotide sequence <SEQ ID 359> is:

	1	ATGANACACA	TACTCCCCCT	GANTGNCGCA	TCCGCACTCT	GCATTTCAAC
	51	CGTTTCGGNN	CATCCTGCCA	GCGAACCGCA	AACCCAAAAC	GAAACCGCTA
15	101	TGACCACGCA	TACCCTCACC	TCAAAATACA	GTTTTGACGA	AACCGTCAGC
	151	CGCCTTGAAA	CGCCATAAA	AAGCAAAGGG	ATGGACATT	TTGCCGTCAT
	201	CGACCATCAG	GAAGCCGCCC	GCCGAAACGG	CTTAACGATG	CAGCCGGCAA
	251	AAGTCATCGT	CTTCGGCACG	CCCAAAGCCG	GTACGCCGCT	GATGGTCAAA
20	301	GACCCCGCCT	TCGCCCTGCA	ACTGCCCTG	CGCGTCNTCG	TTACCGAAAC
	351	GGACGGCAAA	GTACGCGCCG	CCTATACCGA	TACGCGCGCC	CTCATCGCCG
	401	GCAGCCGCAT	CGGTTTCGAC	GAAGTGGCAA	ACACTTTGGC	AAACGCCGAA
	451	AAACTGATAC	AAAAAACCAT	AGGCGAATAA		

This encodes a protein having amino acid sequence <SEQ ID 360>:

	1	MXHILPLXXA	SALCISTASX	HPASEPQTQN	ETAMTTHTLT	SKYSFDETVS
25	51	RLETAIKSKG	MDIFAVIDHQ	EAARRNGLTM	QPAKVIVFGT	PKAGTPLMVK
	101	DPAFALQLPL	RVXVTETDGK	VRAAYTDTRA	LIAGSRIGFD	EVANTLANAE
	151	KLIQKTIGE*				

ORF97a and ORF97-1 show 95.6% identity in 159 aa overlap:

		10	20	30	40	50	60
30	orf97a.pep	MXHILPLXXASALCISTASXHPASEPQTQN	ETAMTTHTLT	SKYSFDETVS	SRLETAIKSKG		
	orf97-1	MKHILPLIAASALCISTASAHASEPSTQN	ETAMTTHTLT	SKYSFDETVS	SRLETAIKSKG		
		10	20	30	40	50	60
35							
	orf97a.pep	MDIFAVIDHQAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVXVTETDGK					
	orf97-1	MDIFAVIDHQAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVTETDGK					
40		70	80	90	100	110	120
	orf97a.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					
45	orf97-1	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					
		130	140	150	160		

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF97 shows 88.1% identity over a 159aa overlap with a predicted ORF (ORF97.ng) from *N. gonorrhoeae*:

50	orf97.pep	MKHILPLIAASALCISTASAHASEPSTQNETAMITHTLISKYSFGXXXXXXXXXAISKSG	60
	orf97ng	MKHILPPIAASAFICISTASAHAPAGKPPTQNETAMTTHTLSKYSFDETVSRLETAISKSG	60
55	orf97.pep	MDIFAVIDHQAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVTETDGK	120
	orf97ng	MDIFAVIDHQAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVTETDGK	120
60	orf97.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVE	159
	orf97ng	VRTAYTDTRALIVGSRISFDEVANTLANAEKLIQKTVE	159

The complete length ORF97ng nucleotide sequence <SEQ ID 361> is predicted to encode a protein having amino acid sequence <SEQ ID 362>:

```

      1 MKHILPPIAA SAFCISTASA HPAGKPPTQN ETAMTTHTLT SKYSFDETVS
      51 RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
5 101 DPAFALQLPL RVLVTETDGK VRTAYTDTRA LIVGSRISFD EVANTLANAE
      151 KLIQKTVGE*

```

Further work revealed the complete nucleotide sequence <SEQ ID 363>:

```

      1 ATGAAACACA TACTCCCcct gatcgccgca TccgcactCT GCATTTCAAC
      51 CGCTTCGGCA CACCCTGCCG GCAAACCGCC CACCCAAAAC GAAACCGCTA
10 101 TGACCACGCA CACCCTCACC TCGAAATACA GTTTTGACGA AACCGTCAGC
      151 CGCCTTGAAA CCGCCATAAA AAGCAAAGGG ATGGACATTT TTGCCGTCAT
      201 CGACCATCAG GAAGCGGCAC GCCGAAACGG CCTGACCATG CAGCCGGCAA
      251 AAGTCATCGT CTTCCGGCAG CCAAGGCCG GTACGCCgct GATGGTCAAA
      301 GACCCCGCCT TCGCCCTGCA ACTGCCCTG CGCGTCCTCG TTACCGAAAC
15 351 GGACGGCAAA GTACGCACCG CCTATACCGA TACGCGCGCC CTCATCGTCG
      401 GCAGCCGCAT CAGTTTCGAC GAAGTGCAA AACTTTGGC AAACGCCGAA
      451 AAACGATAC AAAAAACCGT AGGCGAATAA

```

This corresponds to the amino acid sequence <SEQ ID 364; ORF97ng-1>:

```

      1 MKHILPLIAA SALCISTASA HPAGKPPTQN ETAMTTHTLT SKYSFDETVS
      51 RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
20 101 DPAFALQLPL RVLVTETDGK VRTAYTDTRA LIVGSRISFD EVANTLANAE
      151 KLIQKTVGE*

```

ORF97ng-1 and ORF97-1 show 96.2% identity in 159 aa overlap:

```

25      10      20      30      40      50      60
      orf97-1.pep MKHILPLIAASALCISTASAHPASEPSTQNETAMTTHTLTSKYSFDETVSRLETAIKSKG
      orf97ng-1   MKHILPLIAASALCISTASAHPAKPPTQNETAMTTHTLTSKYSFDETVSRLETAIKSKG
      10      20      30      40      50      60
30      70      80      90      100     110     120
      orf97-1.pep MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVTETDGK
      orf97ng-1   MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVTETDGK
      70      80      90      100     110     120
35      130     140     150     160
      orf97-1.pep VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEX
      orf97ng-1   VRTAYTDTRALIVGSRISFDEVANTLANAEKLIQKTVGEX
40      130     140     150     160

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF97-1 (15.3kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 12A & 12B show, respectively, the results of affinity purification of the GST-fusion and His-fusion proteins. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western Blot (Figure 12C), ELISA (positive result), and FACS analysis (Figure 12D). These experiments confirm that ORF97-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 12E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF97-1.

### Example 43

The following DNA, believed to be complete, sequence was identified in *N.meningitidis* <SEQ ID 365>:

```

5      1  ATGGCTTTTA TTACGCGCTT ATTCAAAGC AGTAAATGGC TGATTGTGCC
      51  GCTGATGCTC CCCGCCTTTC AGAATGTGGC GGCGGAGGGG ATAGATGTGA
     101  GCCGTGCCGA AGCGAGGATA ACCGACGGCG GGCAGCTTTC CATCAGCAGC
     151  CGCTTCCAAA CCGAGCTGCC CGACCAGCTC CAACAGGCGT TGCGCCGGGg
     201  CGTGCCGCTC AACTTTACCT TAAGCTGGCA GCTTTCCGCC CCGATAATCG
     251  CTTCTTATCG GTTTAAATTG GGGCAACTGA TTGGCGATGA CGACaATATT
     301  GACTACAAAC TGAGTTTCCA TCCGCTGACC AaACGCTACC GCGTTACCgT
     351  CGgCGCGTTT TCGACAGACT ACGACACCTT GGATGCGGCA TTGCGCGCGA
     401  CCGGCGCGGT TGCCAACCTG AAAGTCCTGA ACAAAGGCGC GCTGTCCGGT
     451  GCGGAAGCAG GGGAAACCAA GGCGGAAATC CGCCTGACGC TGTCCACTTC
    15   501  AAAACTGCCC AAGCCTTTTC AAATCAATGC ATTGACTTCT CAAAACCTGGC
     551  ATTTGGATTC GGGTTGAAA CCTCTAAACA TCATCGGGAA CAAATAA
  
```

This corresponds to the amino acid sequence <SEQ ID 366; ORF106>:

```

      1  MAFITRLFKS SKWLIVPLML PAFQNVAAEG IDVSRAEARI TDGGQLSISS
     51  RFQTELPDQL QQALRRGVPL NFTLSWQLSA PIIASYREKL GQLIGDDNII
    101  DYKLSFHPLT KRYRVTVGAF STDYDTLDA LRAATGAVANW KVLNKGALSG
     151  AEAGETKAEI RLTLSTSKLP KPFQINALTS QNWHLD SGWK PLNIIGNK*
  
```

Further work revealed the following DNA sequence <SEQ ID 367>:

```

      1  ATGGCTTTTA TTACGCGCTT ATTCAAAGC AGTAAATGGC TGATTGTGCC
     51  GCTGATGCTC CCCGCCTTTC AGAATGTGGC GGCGGAGGGG ATAGATGTGA
    101  GCCGTGCCGA AGCGAGGATA ACCGACGGCG GGCAGCTTTC CATCAGCAGC
     151  CGCTTCCAAA CCGAGCTGCC CGACCAGCTC CAACAGGCGT TGCGCCGGGG
     201  CGTGCCGCTC AACTTTACCT TAAGCTGGCA GCTTTCCGCC CCGATAATCG
     251  CTTCTTATCG GTTTAAATTG GGGCAACTGA TTGGCGATGA CGACaATATT
     301  GACTACAAAC TGAGTTTCCA TCCGCTGACC AACCCTACC GCGTTACCgT
     351  CGGCGCGTTT TCGACAGACT ACGACACCTT GGATGCGGCA TTGCGCGCGA
     401  CCGGCGCGGT TGCCAACCTG AAAGTCCTGA ACAAAGGCGC GCTGTCCGGT
     451  GCGGAAGCAG GGGAAACCAA GGCGGAAATC CGCCTGACGC TGTCCACTTC
     501  AAAACTGCCC AAGCCTTTTC AAATCAATGC ATTGACTTCT CAAAACCTGGC
     551  ATTTGGATTC GGGTTGAAA CCTCTAAACA TCATCGGGAA CAAATAA
  
```

This corresponds to the amino acid sequence <SEQ ID 368; ORF106-1>:

```

      1  MAFITRLFKS SKWLIVPLML PAFQNVAAEG IDVSRAEARI TDGGQLSISS
     51  RFQTELPDQL QQALRRGVPL NFTLSWQLSA PIIASYREKL GQLIGDDNII
    101  DYKLSFHPLT NRYRVTVGAF STDYDTLDA LRAATGAVANW KVLNKGALSG
     151  AEAGETKAEI RLTLSTSKLP KPFQINALTS QNWHLD SGWK PLNIIGNK*
  
```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF106 shows 87.4% identity over a 199aa overlap with an ORF (ORF106a) from strain A of *N.meningitidis*:

```

45      orf106.pep      10      20      30      40      50      59
      MAFITRLFKSSK-WLIVPLMLPAFQNVAAEGIDVSRAEARITDGGQLSISSRFQTELPDQ
      ||||| ||| : || : : ||||| ||||| : ||||| ||||| |||||
      orf106a          MAFITRLFKSIKQWLVLPLSLVLPDAAEGIDVSRAEARIXDGGQLSXXSRFQTELPDQ
      10      20      30      40      50      60

50      orf106.pep      60      70      80      90      100      110      119
      LQQALRRGVPLNFTLSWQLSAPIIASYREKLGQLIGDDNIDYKLSFHPLTKRYRVTGVA
      || | ||| || || ||||| ||||| ||||| ||||| : |||||
  
```



-235-

```

orfl06a      LQXAXXRGVXLNXTLXWQLSAPIIASYRFXLGQLIGDDDXIDYKLSFHPLTNRYRVTVGA
              70      80      90      100     110     120

5  orfl06.pep 120      130      140      150      160      170      179
    orfl06a   FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSTSKLPKPFQINALT
              ||| |||||
    orfl06a   FSTXYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSTSKLPKPFQINALT
              130      140      150      160      170      180

10 orfl06.pep 180      190      199
    orfl06a   SQNWHLD SGWKPLNIIGNKX
              |||||
    orfl06a   SQNWHLD SGWKPLNIIGNKX
              190      200

```

15 Due to the K→N substitution at residue 111, the homology between ORF106a and ORF106-1 is 87.9% over the same 199 aa overlap.

The complete length ORF106a nucleotide sequence <SEQ ID 369> is:

```

20 1 ATGGCTTTTA TTACGCGCTT ATTCAAAAGC ATTAAACAAT GGCTTGTGCT
    51 GCTGCCGATG CTTTCCGTTT TGCCGGACGC GCGGCGGAG GGGATAGATG
    101 TGAGCCGCGC CGAAGCGAGG ATAANCGACG GCGGGCAGCT TTCCATNAGN
    151 AGCCGCTTCC AAACCGAGCT GCCCGACCAG CTCCAANNNG CGNNGNGCCG
    201 GGGCGTGNCG CTCAACTNTA CCTTAAGNTG GCAGCTTTC GCGCCGATAA
    251 TCGCTTCTTA TCGGTTTNAA TTGGGGCAAC TGATTGGCGA TGACGACNAT
    301 ATTGACTACA AACTGAGTTT CCATCCGCTG ACCAACCGCT ACCGCGTTAC
    351 CGTCGGCGCG TTTTCGACAG ANTACGACAC CTTGGATGCG GCATTGCGCG
    401 CGACCGGCGC GGTTCGCAAC TGGAAAGTCC TGAACAAAGG CGCGCTGTCC
    451 GGTGCGGAAG CAGGGGAAAC CAAGGCGGAA ATCCGCCTGA CGCTGTCCAC
    501 TTCAAAACTG CCCAAGCCTT TTCAAATCAA TGCATTGACT TCTCAAAACT
    551 GGCATTGGA TTCGGGTGG AAACCTCTAA ACATCATCGG GAACAAATAA

```

30 This encodes a protein having amino acid sequence <SEQ ID 370>:

```

35 1 MAFITRLFKS IKQWLVLPM LSVLPDAAAE GIDVSRAEAR IXDGGQLSXX
    51 SRFQTELPDQ LQXAXXRGVX LNXTLXWQLS APIIASYRFX LGQLIGDDDX
    101 IDYKLSFHPL TNRYRVTVGA FSTXYDTLDA ALRATGAVAN WKVLNKGALS
    151 GAEAGETKAE IRLTLSTSKL PKPFQINALT SQNWHLD SGW KPLNIIGNK*

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF106 shows 90.5% identity over a 199aa overlap with a predicted ORF (ORF106.ng) from *N. gonorrhoeae*:

```

40 orfl06.pep MAFITRLFKSSK-WLIVPLMLPAFQNVAAEGIDVSRAEARITDGGQLSISRRFQTELPDQ 59
    orfl06ng  MAFITRLFKSIKQWLVLPLISVLPDAAAEGIAATRAEARITDGGRLSISRRFQTELPDQ 60

    orfl06.pep LQQALRRGVPLNFTLSWQLSAPIIASYRFXLGQLIGDDDNIDYKLSFHPLTKRYRVTVGA 119
    orfl06ng  LQQALRRGVPLNFTLSWQLSAPTIASYRFXLGQLIGDDDNIDYKLSFHPLTNRYRVTVGA 120

45 orfl06.pep FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSTSKLPKPFQINALT 179
    orfl06ng  FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSTSKLPKPFQINALT 180

50 orfl06.pep SQNWHLD SGWKPLNIIGNK 198
    orfl06ng  SQNWHLD SGWKPLNIIGNK 199

```

55 Due to the K→N substitution at residue 111, the homology between ORF106ng and ORF106-1 is 91.0% over the same 199 aa overlap.

The complete length ORF106ng nucleotide sequence <SEQ ID 371> is:

```

1   ATGGCTTTTA TTACGCGCTT ATTCAAAAGC ATTAAACAAT GGCTTGTGCT
51  GTTGCCGATA CTCTCCGTTT TGCCGGACGC GCGGCGGAG GGCATTGCCG
101 CGACCCGCGC CGAAGCGAGG ATAACCGACG GCGGGCGGCT TTCCATCAGC
5   151 AGCCGCTTCC AAACCGAGCT GCCCGACCAG CTCCAACAGG CGTTGCGCCG
201 GGGCGTACCG CTCAACTTTA CCTTAAGCTG GCAGCTTCC GCCCGACAA
251 TCGCTTCTTA TCGGTTTAAA TTGGGGCAAC TGATTGGCGA TGACGACAAT
301 ATTGACTACA AACTAAGTTT CCATCCGCTG ACCAACCGCT ACCGCGTTAC
10 351 CGTCGGCGCA TTTTCCACCG ATTACGACAC TTTGGATGCG GCATTGCGCG
401 CGACCGGCGC GGTTGCCAAC TGGAAAGTCC TGAACAAAGG CGCGTTGTCC
451 GGTGCGGAAG CAGGGGAAAC CAAGGCGGAA ATCCGCCTGA CGCTGTCCAC
501 TTCAAAACTG CCCAAGCCTT TCCAAATCAA CGCATTGACT TCTCAAACT
551 GGCATTGGA TTGCGGTTGG AAACCTCTAA ACATCATCGG GAACAAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 372>:

```

15 1   MAFITRLFKS IKQWLVL LPI LSVLPDAAAE GIAATRAEAR ITDGGRLSIS
51 51  SRFQTELPDQ LQQALRRGVP LNFTLSWQLS APTIASYRFK LGQLIGDDDN
101 101 IDYKLSFHPL TNRYRVTGVA FSTDYDTLDA ALRATGAVAN WKVLNKGALS
151 151 GAEAGETKAE IRLTLSTSKL PKPFQINALT SQNWHLDSGW KPLNIIGNK*

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF106-1 (18kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 13A shows the results of affinity purification of the His-fusion protein, and Figure 13B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 13C) These experiments confirm that ORF106-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 44

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 373>:

```

1   ATGGACACAA AAGAAATCCT CGG.TACGCG GcAGGcTCGA TCGGCAGCGC
51 51  GGTTTTAGCC GTCATCATCc TGCCGCTGCT GTCGTGGTAT TTCCCCGCCG
101 ACGACATCGG GCGCATCGTG CTGATGCAGA CCGCGGCGGG GCTgACGGTG
151 TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
35 201 CACCGCCGAC AAAGACAcCT TGTTCAAAAC CCTGTTCTCTG CCGCCGCTGC
251 TGTCTGCCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
301 TCTGAAATCC TGTTTTCACT CGACGATGCC gCCGCCGGCa TCGGGCTGGT
351 GCTGTTTGAA CtGAGCTTCC TGCCCATCCG cTTTCTCTTA CTGGTTTTC
40 401 GTATGGAAGG ACGCGCCcTT GCCTTTTCGT CCGCGCAACT CGTGCCCAAG
451 CTCGCCATCC TGCTGCTG.T GCCGCTGACG GTCGGGCTGC TGCACTTTCC
501 AGCGAACACC GCCGTCTGA CCGCCGTTA CGCGCTGGCA AACCTTGCCG
551 CCGCCGCCTT TTTGCTGTTT CAAAACCGAT GCCGTCTGAA GGCCGTCCGG
601 CACGCACCGT TTTGCCCCGC CGTCTGCAC CGGGG.TGC GCTACGGCAT
651 ACCGATCGCA CTGAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCCGACC
45 701 GTTGTTCCTT GAAAAAATAT GCCGGCCTGG AACAGCTCGG CGTTTATTTC
751 ATGGGTATTT CGTTCGGCGG GGCGGCATTA TTGTTCCAAA GCATCTTTTC
801 AACGGTCTGG ACACCGTATA TTTTCCGCGC AATCGAAGAA AACGCCCCGC
851 CCGCTCGCCT CTCGGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCTTCC
901 GCCCTCTGC. TGACCGGCAT TTTCTCGCCC CTTGCCTCCC TCCTGCTGCC
50 951 GGAAACTAC GCCCGCTGCC GGTTTATCGT CGTATCGTGT ATG.TGCCGC

```

-237-

```

1001 CGCTGTTTTG CACGCTGGCG GAAATCAGCG GCATCGGTTT GAACGTCGTT
1051 CGCAAAACGC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA
1101 CTGCTGCTG CTGGGGCTTG ACCGTGCCGT ACCGGCGAGG CCGCC.GGCG
1151 CGGCGGTTGC CTGTGCCGCC TCATTCTGGC TGTTTTTTGC CTTCAAGACC
1201 GAAAGCTCyT GCCGCTGTG GCAGCCGTC AACGCCTGC CGCTTTATCT
1251 GCACACATTG TTCTGCCTGA CCTCCTCGGC GGCCTACACC TGCTTCGGCA
1301 CGCCGGCAAA CTATCCCCTG TTTGCCGGCG TATGGGCGGC ATATCTGGCA
1351 GGCTGCATCC TGCGCCACCG GAAAGATTG CACAACTGT TTCATTATTT
1401 GAAAAACAA GGTTCCTCAT TATGA

```

10 This corresponds to the amino acid sequence <SEQ ID 374; ORF10>:

```

1 MDTKEILXYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
51 SVLCLGLDQA YVREYYATAD KDTLFTKTLFL PPLLSAAAIA ALLLSRPSLP
101 SEILFSLDDA AAGIGLVLF LSLFLPIRFL LVLRLMEGRAL AFSSAQLVPK
151 LAILLXPLT VGLLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKAVR
201 HAPFSPAVLH RGXRYGIPIA LSSIAYWGLA SADRLFLKKY AGLEQLGVYS
251 MGISFGGAAL LFQSFSTVW TPYIFRAIEE NAPPARLSAT AESAAALLAS
301 ALCXTGIFSP LASLLLPENY AAVRFIVVSC MXPPLFCTLA EISGIGLNVV
351 RKTRPIALAT LGALANLLL LGLDRAVPAR PXGAAVACAA SFWLFFAFKT
401 ESSCRLWQPL KRLPLYLHTL FCLTSSAAYT CFGTPANYPL FAGVWAAVLA
20 451 GCILRHRKDL HKLFHYLKKQ GFPL*

```

Further sequence analysis revealed the complete DNA sequence<SEQ ID 375> to be:

```

1 ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC
51 GGTTTTAGCC GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCCCCCGCG
25 101 ACGACATCGG GCGCATCGTG CTGATGCAGA CGGCGGCGGG GCTGACGGTG
151 TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
201 CACCGCCGAC AAAGACACCT TGTTCAAAAC CCTGTTCTCTG CCGCCGCTGC
251 TGCTGCGCGC CGCGATAGCC GCCCTGCTGC TTTCCCCGCC GTCCCTGCCG
301 TCTGAAATCC TGTTTCACT CGACGATGCC GCCGCGGCA TCGGGCTGGT
351 GCTGTTTGAA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTTGC
30 401 GTATGGAAGG ACGCGCCCTT GCCTTTTCTG CCGCGCAACT CGTGCCCAAG
451 CTCGCCATCC TGCTGCTGCT GCCGCTGACG GTCGGGCTGC TGCATTTC
501 AGCGAACACC GCCGCTCTGA CCGCGTTTA CGCGCTGGCA AACCTTGCCG
551 CGGCCGCCTT TTGCTGTTT CAAAACCGAT GCCGCTGAA GGCCGTCGG
601 CACGCACCGT TTTCGCGCGC CGTCTGCAC CGGGGCTGC GCTACGGCAT
35 651 ACCGATCGCA CTGAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCCGACC
701 GTTTGTTCC TAAAAAATAT GCCGGCCTGG AACAGCTCGG CGTTTATTCG
751 ATGGGTATTT CGTTCGGCGG GGCGGCATTA TTGTCCAAA GCATCTTTTC
801 AACGGTCTGG ACACCGTATA TTTCCGCGC AATCGAAGAA AACGCCCGC
851 CCGCCGCCT CTCGGAACG GCAGAATCCG CCGCCGCCCT GCTTGCTCC
40 901 GCCCTCTGCC TGACCGGCAT TTTCTCGCCC CTTGCCTCCC TCCTGTGCC
951 GGAAACTAC GCCGCCGTCC GGTATATCGT CGTATCGTGT ATGCTGCCG
1001 CGCTGTTTTG CACGCTGGCG GAAATCAGCG GCATCGGTTT GAACGTCGTC
1051 CGCAAAACGC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA
1101 CTGCTGCTG CTGGGGCTTG CCGTGCCGTC CGGCGGCGCG CGCGGCGCGG
45 1151 CGGTTGCCTG TGCCGCTCA TTCTGGCTGT TTTTGCCTT CAAGACCGAA
1201 AGCTCCTGCC GCCTGTGGCA GCCGCTCAA CGCCTGCCG TTTATCTGCA
1251 CACATTGTTG TGCCTGACCT CCTCGGCGGC CTACACCTGC TTCGGCACGC
1301 CGGCAAACTA TCCCCTGTTT GCCGGCGTAT GGGCGGCATA TCTGGCAGG
1351 TGCATCCTGC GCCACCGGAA AGATTGTCAC AAAGTGTTC ATTATTGAA
50 1401 AAAACAAGGT TTCCATTAT GA

```

This corresponds to the amino acid sequence <SEQ ID 376; ORF10-1>:

```

1 MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
51 SVLCLGLDQA YVREYYATAD KDTLFTKTLFL PPLLSAAAIA ALLLSRPSLP
55 101 SEILFSLDDA AAGIGLVLF LSLFLPIRFL LVLRLMEGRAL AFSSAQLVPK
151 LAILLXPLT VGLLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKAVR
201 HAPFSPAVLH RGLRYGIPIA LSSIAYWGLA SADRLFLKKY AGLEQLGVYS
251 MGISFGGAAL LFQSFSTVW TPYIFRAIEE NAPPARLSAT AESAAALLAS
301 ALCLTGIFSP LASLLLPENY AAVRFIVVSC MLPPLFCTLA EISGIGLNVV
351 RKTRPIALAT LGALANLLL LGLAVPSGGA RGAAVACAAS FWLFFAFKTE
60 401 SSCRLWQPLK RLPLYLHTLF CLTSSAAYTC FGTPANYPLF AGVWAAVLAG
451 CILRHRKDLH KLFHYLKKQG FPL*

```

Computer analysis of this amino acid sequence gave the following results:



-239-

		310	320	330	340	350	360
	orf10.pep	ALCXTGIFSP	LLPENYA	AVRFIVV	SCMXPPL	FCTLA	EISGIGLNVVRKTRPIALAT
5	orf10a	ALCLTGIFSP	LLPENYA	AVRFIVV	SCMLPPL	FCTLVE	ISGIGLNVVRKTRPIALAT
		310	320	330	340	350	360
		370	380	390	400	410	419
	orf10.pep	LGALAANL	LLGLDRA	VPAR-PX	GAAVACA	SEFWLFF	AFKTESSCRLWQPLKRLPLYLHT
10	orf10a	LGALAANL	LLGL--	AVPSGG	GARGAA	VACAAS	FWLFFVFKTESSCRLWQPLKRLPLYMHT
		370	380	390	400	410	
		420	430	440	450	460	470
	orf10.pep	LFCLTSSA	AYTCFG	TPANYP	LEFAGV	WAAYLA	GCILRHRKDLHKLHFHYLKKQGFPLX
15	orf10a	LFCLASSA	AYTCFG	TPANYP	LEFAGV	WAVYLA	GCILRHRKDLHKLHFHYLKKQGFPLX
		420	430	440	450	460	470

The complete length ORF10a nucleotide sequence <SEQ ID 377> is:

	1	ATGGACACAA	AAGAAATCCT	CGGCTACGCG	GCAGGCTCGA	TCGGCAGCGC
20	51	GGTTTtagcc	GTCATCATCC	TGCCGCTGCT	GTCGTGGTAT	TTCCCTGCCG
	101	ACGACATCGG	ACGCATCGTG	CTGATGCAGA	CGGCGGCGGG	GCTGACGGTG
	151	TCGGTGTTGT	GCCTCGGGCT	GGATCAGGCA	TACGTCCGCG	AATACTATGC
	201	CGCCGCCGAC	AAAGACACTT	TGTTCAAAAC	CCTGTTCCCTG	CCGCCGCTGC
25	251	TGCTGCGCGC	CGCGATAGCC	GCCCTGCTGC	TTTCCCGCCC	ATCCCTGCCG
	301	TCTGAAATCC	TGTTTTCGCT	CGACGATGCC	GCCGCCGGCA	TCGGGCTGGT
	351	GCTGTTTGAA	CTGAGCTTCC	TGCCCATCCG	CTTTCTCTTA	CTGGTTTTGC
	401	GTATGGAAGG	ACGCGCCCTT	GCCTTTTCGT	CCGCGCAACT	CGTGTCCAAG
	451	CTCGCCATCC	TGCTGCTGCT	GCCGCTGACG	GTCCGGGCTGC	TGCACTTTCC
30	501	GGCGAACACC	GCCGTCTGA	CCGCCGTTTA	CGCGCTGGCA	AACCTTGCCG
	551	CCGCCGCCTT	TTTGCTGTTT	CAAAACCGAT	GCCGTCTGAA	GGCCGTCCGG
	601	CGCGCACCGT	TTTCATCCGC	CGTCTGTCAT	CGCGGCCTGC	GCTACGGCAT
	651	ACCGATCGCA	CTAAGCAGCA	TCGCCTATTG	GGGGCTGGCA	TCCGCCGACC
	701	GTTTGTTCCT	GAAAAATAT	GCCGGCCTAG	AACAGCTCGG	CGTTTATTCG
35	751	ATGGGTATTT	CGTTCGGCGG	AGCGGCATTA	TTGTTCCAAA	GCATCTTTTC
	801	AACGGTCTGG	ACACCGTATA	TTTTCCGCGC	AATCGAAGCA	AACGCCCCGC
	851	CCGCCCGCCT	CTCGGCAACG	GCAGAATCCG	CCGCCGCCCT	GCTTGCCCTC
	901	GCCCTCTGCC	TGACCGGCAT	TTTCTCGCCC	CTCGCCTCCC	TCCTGCTGCC
	951	GGAAACTAC	GCCGCCGTCC	GGTTTATCGT	CGTATCGTGT	ATGCTGCCCTC
40	1001	CGCTGTTTTG	CACGCTGGTA	GAAATCAGCG	GCATCGGTTT	GAACGTCGTC
	1051	CGAAAAACAC	GCCCGATCGC	GCTCGCCACC	TTGGGCGCGC	TGGCGGCAAA
	1101	CCTGCTGCTG	CTGGGGCTTG	CCGTACCGTC	CGGCGGCGCG	CGCGGCGCGG
	1151	CGGTGCTGCTG	TGCCGCCTCA	TTTTGGCTGT	TTTTTGTTTT	CAAGACCGAA
	1201	AGCTCCTGCC	GCCTGTGGCA	GCCGCTCAAA	CGCCTGCCGC	TTTATATGCA
45	1251	CACATTGTTT	TGCTTGGCCT	CCTCGGCGGC	CTACACCTGC	TTCGGCCTC
	1301	CGGCAAACTA	CCCCCTGTTT	GCCGGCGTAT	GGGCGGTATA	TCTGGCAGGC
	1351	TGCATCCTGC	GCCACCGGAA	AGATTGTCAC	AAACTGTTTC	ATTATTGAA
	1401	AAAACAAGGT	TTCCCATAT	GA		

This encodes a protein having amino acid sequence <SEQ ID 378>:

	1	MDTKEILGYA	AGSIGSAVLA	VIILPLLSWY	FPADDIGRIV	LMQTAAGLTV
50	51	SVLCLGLDQA	YVREYYAAAD	KDTLFLKTLFL	PPLLSAAAIA	ALLLSRPSLP
	101	SEILFSLDDA	AAGIGLVLFE	LSFLPIRFL	LVLRMGRAL	AFSSAQLVSK
	151	LAILLLLPLT	VGLLHFPANT	AVLTAVYALA	NLAAAFLFF	QNRCLKAVR
	201	RAPFSSAVLH	RGLRYGIPIA	LSSIAYWGLA	SADRLFLKKY	AGLEQLGVYS
55	251	MGISFGGAAL	LFQSIFSTVW	TPYIFRAIEA	NAPPARLSAT	AESAAALLAS
	301	ALCLTGIFSP	LASLLLPENY	AAVRFIVVSC	MLPPLFCTIV	EISGIGLNVV
	351	RKTRPIALAT	LGALAANLLL	LGLAVPSGGA	RGAAVACAAS	FWLFFVFKTE
	401	SSCRLWQPLK	RLPLYMHTLF	CLASSAAYTC	FGTPANYPLF	AGVWAVYLAG
	451	CILRHRKDLH	KLFHYLKKQG	FPL*		

ORF10a and ORF10-1 show 95.4% identity in 475 aa overlap:

		10	20	30	40	50	60
60	orf10-1.pep	MDTKEILXYA	AGSIGSAVLA	VIILPLLSWY	FPADDIGRIV	LMQTAAGLTV	SVLCLGLDQA
	orf10a	MDTKEILGYA	AGSIGSAVLA	VIILPLLSWY	FPADDIGRIV	LMQTAAGLTV	SVLCLGLDQA
		10	20	30	40	50	60

65

-240-

		70	80	90	100	110	120
	orf10-1.pep	YVREYYATADKDTL	FKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLF				
	orf10a	YVREYYAAADKDTL	FKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLF				
5		70	80	90	100	110	120
	orf10-1.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLA	ILLXPLTVGLLHFPANTAVLTAVYALA				
	orf10a	LSFLPIRFLLLVLRMEGRALAFSSAQLVSKLA	ILLXPLTVGLLHFPANTAVLTAVYALA				
10		130	140	150	160	170	180
	orf10-1.pep	NLAAAFLLFQNRCLKAVRHAPFSPAVLHRGXRYGIPIALSSIA	YWG	LASADRLFLK			
	orf10a	NLAAAFLLFQNRCLKAVRRAPFSSAVLHRGLRYGIPIALSSIA	YWG	LASADRLFLK			
15		190	200	210	220	230	240
	orf10-1.pep	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS					
	orf10a	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEANAPPARLSATAESAAALLAS					
20		250	260	270	280	290	300
	orf10-1.pep	ALCXTGIFSPLASLLLPENYA	AVRFIVVSCMPPLFCTLA	EISGIGLN	VVRKTRPIALAT		
	orf10a	ALCLTGIFSPLASLLLPENYA	AVRFIVVSCMLPPLFCTLVEISGIGLN	VVRKTRPIALAT			
25		310	320	330	340	350	360
	orf10-1.pep	LGALAANLLLLGLDRAV	PAR-PXGA	AVACAASF	WLF	FAFKTESSCRLWQPLKRLPLYLHT	
	orf10a	LGALAANLLLLGL--AVPSGG	ARGA	AVACAASF	WLF	FVFKTESSCRLWQPLKRLPLYMHT	
30		370	380	390	400	410	419
	orf10-1.pep	LFCLTSSAAYTCFGTPANYPLFAGVWAAYLAGCILRHRKDLHKL	FHYLKKQGFPLX				
	orf10a	LFCLASSAAYTCFGTPANYPLFAGVWAVYLAGCILRHRKDLHKL	FHYLKKQGFPLX				
35		420	430	440	450	460	470
	orf10-1.pep	ALCXTGIFSPLASLLLPENYA	AVRFTV	VSCMLPFYTLTEISGIGLN	VVRKTRPIALAT		
	orf10a	ALCXTGIFSPLASLLLPENYA	AVRFIVVSCMPPLFCTLA	EISGIGLN	VVRKTRPIALAT		
40		420	430	440	450	460	470

Homology with a predicted ORF from *N.gonorrhoeae*ORF10 shows 94.1% identity over a 475aa overlap with a predicted ORF (ORF10.ng) from *N.*45 *gonorrhoeae*:

	orf10ng.pep	MDTKEILGYAAGSIGSAVLAVIILPPLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA	60
	orf10nm	MDTKEILXYAAGSIGSAVLAVIILPPLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA	60
50	orf10ng.pep	YVREYYAAADKDTL	FKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLF 120
	orf10nm	YVREYYATADKDTL	FKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLF 120
55	orf10ng.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLA	ILLXPLTVGLLHFPANTSVLTAVYALA 180
	orf10nm	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLA	ILLXPLTVGLLHFPANTAVLTAVYALA 180
60	orf10ng.pep	NLAAAFLLFQNRCLKAVRRAPFSPAVLHRGLRYGIPIALSSLAYWG	LASADRLFLK 240
	orf10nm	NLAAAFLLFQNRCLKAVRHAPFSPAVLHRGXRYGIPIALSSIA	YWG LASADRLFLK 240
65	orf10ng.pep	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENATPARLSATAESAAALLAS	300
	orf10nm	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS	300
	orf10ng.pep	ALCXTGIFSPLASLLLPENYA	AVRFTV VSCMLPFYTLTEISGIGLN VVRKTRPIALAT 360
	orf10nm	ALCXTGIFSPLASLLLPENYA	AVRFIVVSCMPPLFCTLA EISGIGLN VVRKTRPIALAT 360

-241-

		370	380	390	400	410
5	orf10ng.pep	LGALAANLLLLGL--AVPSGGTRGA	AVACAASFWLFFVKTESSCRLWQPLKRLPLYMHT			
	orf10nm	LGALAANLLLLGLDRAVPAR-PXGA	AVACAASFWLFFAFKTESSCRLWQPLKRLPLYLHT			
		370	380	390	400	410
10	orf10ng.pep	LFCLASSAAYTCFGTPANYPLFAGV	WAAYLAGCILRHRKNLHKLFHYLKKQGFPLX			
	orf10nm	LFCLTSSAAYTCFGTPANYPLFAGV	WAAYLAGCILRHRKDLHKLFHYLKKQGFPLX			
		420	430	440	450	460
		420	430	440	450	460

The complete length ORF10ng nucleotide sequence <SEQ ID 379> is:

	1	ATGGACACAA	AAGAAATCCT	CGGCTACGCG	GCAGGCTCGA	TCGGCAGCGC
15	51	GGTTTTAGCC	GTCATCATCC	TGCCGCTGCT	GTCGTGGTAT	TTCcccgCCG
	101	ACGACATCGG	GCGCATCGTG	CTGATGCAGA	CGGCGGCGGG	ACTGACGGTG
	151	TCGGTATTGT	GCCTCGGGCT	GGATCAGGCA	TACGTCCGCG	AATACTATGC
	201	CGCCGCCGAC	AAAGACACTT	TGTTCAAAAC	CCTGTTCTCT	CCGCCGCTGC
20	251	TGTTTCCGC	CGCGATAGCC	GCCCTGCTGC	TTTCCCGCCC	GTCCCTGCCG
	301	TCTGAAATCC	TGTTTTCGCT	CGACGATGCC	GCCGCCGCA	TCGGGCTGGT
	351	GCTGTTTGAA	CTGAGCTTCC	TGCCCATCCG	CTTTCTCTTA	CTGGTTTTGC
	401	GTATGGAAGG	GCGCGCCCTT	GCCTTTTCGT	CCGCGCAACT	CGTGCCCAAA
	451	CTCGCCATTC	TGCTGCTGTT	GCCGCTGACG	GTCCGGCTGC	TGCACTTTCC
25	501	GGCGAACACC	TCCGTCTGA	CCGCCGTTTA	CGCGCTGGCA	AACCTTGCCG
	551	CCGCCGCCTT	TTTGCTGTTT	CAAACCGAT	GCCGTCTGAA	GGCCGTCCGG
	601	CGCGCGCCGT	TTTCGCCCGC	CGTCTGCAC	CGGGGGCTGC	GCTACGGCAT
	651	ACCGCTCGCA	CTGAGCAGCC	TTGCCTATTG	GGGGCTGCA	TCCGCCGACC
	701	GTTTGTTCTT	GAAAAAATAT	GCGGGCCTGG	AACAGCTCGG	CGTTTATTCG
30	751	ATGGGTATTT	CGTTCGGCGC	GGCGGCATTA	TTGCTCCAAA	GCATCTTTTC
	801	AACGGTCTGG	ACACCGTATA	TTTTCCGTGC	AATCGAAGAA	AACGCCACGC
	851	CGCCCCGCT	CTCGGCAACG	GCAGAATCCG	CCGCCGCCCT	GCTTGCCCTCC
	901	GCCTCTGCC	TGACCGGAAT	TTTCTCGCCC	CTCGCCTCCC	TCCTGTGCC
	951	GGAAACTAC	GCCGCCGTCC	GGTTTACCGT	CGTATCGTGT	ATGCTGccgc
35	1001	cgctGTTTTA	CACGCTGACC	GAAATCAGCG	GCATCGGTTT	GAACGTGCTC
	1051	CGCAAAACGC	GTCCGATCGC	GCTTGCCACC	TTGGGCGCGC	TGGCGGCAAA
	1101	CCTGCTGCTG	CTGGGGCTTG	CCGTACCGTC	CGGCGGCACG	CGCGGCGCGG
	1151	CGTTTGCTG	TGCCCGCTCA	TTCTGGTTGT	TTTTTGTTTT	CAAGACAGAA
	1201	AGCTCCTGCC	GCCTGTGGCA	GCCGCTCAA	CGCCTGCCGC	TTTATATGCA
40	1251	CACATTGTTT	TGCCTgGCCT	CCTCGGCGGC	CTACACCTGC	TTCGGCACAC
	1301	CGGCAACTA	CCCcctgttt	gccggcgctAT	GGGCGGCATA	TCTGGCAGGC
	1351	TGCATCCTGC	GCCACCGGAA	AAATTTGCAC	AACTGTTTC	ATTATTGAA
	1401	AAAACAAGGT	TTCCATTAT	GA		

This encodes a protein having amino acid sequence <SEQ ID 380>:

	1	MDTKEILGYA	AGSIGSAVLA	VIILPLLSWY	FPADDIGRIV	LMQTAAGLTV
45	51	SVLCLGLDQA	YVREYYAAAD	KDTLFKTLFL	PPLLFSAIA	ALLSRPSLP
	101	SEILFSLDDA	AAGIGLVLE	LSFLPIRFL	LVLMEGRAL	AFSSAQLVPK
	151	LAILLLLPLT	VGLLHFPANT	SVLTAVYALA	NLAAAFLLF	QNRCLKAVR
	201	RAPFSPAVLH	RGLRYGIPLA	LSSLAYWGLA	SADRLFLKKY	AGLEQLGVYS
50	251	MGISFGGAAL	LLQSIFSTVW	TPYIFRAIEE	NATPARLSAT	AESAAALLAS
	301	ALCLTGIFSP	LASLLLPENY	AAVRFTVVSC	MLPPLFYTLT	EISGIGLNVV
	351	RKTRPIALAT	LGALAANLLL	LGLAVPSGGT	RGAAVACAAS	FWLFFVEKTE
	401	SSCRLWQPLK	RLPLYMHTLF	CLASSAAYTC	FGTPANYPLF	AGVWAAYLAG
	451	CILRHRKNLH	KLFHYLKKQG	FPL*		

ORF10ng and ORF10-1 show 96.4% identity in 473 aa overlap:

55		10	20	30	40	50	60
	orf10-1.pep	MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA					
	orf10ng-1	MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA					
60		10	20	30	40	50	60
	orf10-1.pep	YVREYYATADKDTLFKTLFLPPLLSAAIAALLSRPSLPSEILFSLDDAAAGIGLVLE					
	orf10ng-1	YVREYYAAADKDTLFKTLFLPPLLSAAIAALLSRPSLPSEILFSLDDAAAGIGLVLE					
65		70	80	90	100	110	120

-242-

		130	140	150	160	170	180
	orf10-1.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLPLTVGLLHFPANTAVLTAVYALA					
5	orf10ng-1	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLPLTVGLLHFPANTSVLTAVALA					
		130	140	150	160	170	180
	orf10-1.pep	NLAAAFLLFQNRCLKAVRHAPFSPAVLHRGLRYGIPIALSSSIAYWGLASADRLFLKKY					
10	orf10ng-1	NLAAAFLLFQNRCLKAVRRAPFSPAVLHRGLRYGIPLALSSLAYWGLASADRLFLKKY					
		190	200	210	220	230	240
	orf10-1.pep	AGLEQLGVYSMGISFGGAALLQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS					
15	orf10ng-1	AGLEQLGVYSMGISFGGAALLQSIFSTVWTPYIFRAIEENATPARLSATAESAAALLAS					
		250	260	270	280	290	300
	orf10-1.pep	ALCLTGIFSPLASLLLPENYAARFIVVSCMLPPLFCTLAELISGIGLNVVRKTRPIALAT					
20	orf10ng-1	ALCLTGIFSPLASLLLPENYAARFIVVSCMLPPLFYTLTEISGIGLNVVRKTRPIALAT					
		310	320	330	340	350	360
	orf10-1.pep	LGALAANLLLGLAVPSGGARGA AVACAASFVWFFFAFKTESSCRLWQPLKRLPLYLHTLF					
25	orf10ng-1	LGALAANLLLGLAVPSGGTRGA AVACAASFVWFFVFKTESSCRLWQPLKRLPLYMHTLF					
		370	380	390	400	410	420
	orf10-1.pep	CLTSSAAYTCFGTPANYPLFAGVWAAYLAGCILRHRKDLHKLFHYLKKQGFP LX					
30	orf10ng-1	CLASSAAYTCFGTPANYPLFAGVWAAYLAGCILRHRKNLHKLFHYLKKQGFP LX					
		430	440	450	460	470	

Based on this analysis, including the presence of a putative leader peptide and several  
 40 transmembrane segments and the presence of a leucine-zipper motif (4 Leu residues spaced by 6  
 aa, shown in bold), it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and  
 their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 45

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 381>:

45	1..ATCCTGAAAC CGCATAACCA GCTTAAGGAA GACATCCAAC CTGATCCGGC
	51 CGATCAAAAC GCCTTGTCGG AACC GGATGC TGCGACAGAG GCAGAGCACT
	101 CGGATGCGGA AAATGCTGCC GACAAGCAGC CCGTTGCCGA TAAAGCCGAC
	151 GAGGTTGAAG AAAAGGCGGG CGAGCCGGAA CGGGAAGAGC CGGACGGACA
	201 GGCAGTGCGT AAGAAAGCGC TGACGGAAGA GCGTGAACAA ACCGTCAGGG
50	251 AAAAAGCGCA GAAGAAAGAT GCCGAAACGG TTAAATACA AGCGGTAAAA
	301 CCGTCTAAAG AAACAGAGAA AAAAGCTTCA AAAGAAGAGA AAAAGGCGGC
	351 GAAGGAAAA GTTGCACCCA AACCAACCC GGAACAAATC CTCAACAGCG
	401 GCAGCATCGA AAAGCGCGC AGTGCCGCG CCAAAGAAGT GCAGAAATG
	451 AA.AACGTCC GACAAGGCGG AAGC.AACGC ATTATCTGCA AATGGGCGCG
55	501 TATGCCGACC GTCAGAGCGC GGAAGGGCAG CGTGCCAAAC TGGCAATCTT
	551 GGGCATATCT TCCAAGGTGG TCGGTTATCA GGCGGGACAT AAAACGCTTT
	601 ACCGGGTGCA AAGCGGCAAT ATGTCTGCCG ATGCGGTGA

This corresponds to the amino acid sequence <SEQ ID 382; ORF65>:

60	1..ILKPHNQLKE DIQDPADQN ALSEPDAATE AEQSDAENAA DKQPVADKAD
	51 EEEKAGEPE REEPDQAVR KKALTEEREQ TVREKAQKKD AETVKIQAVK



-243-

101 PSKETEEKAS KEEKKAAKEK VAPKPTPEQI LNSGSIEXAR SAAAKEVQKM  
 151 XNVRQGGXSR IICKWARMPT VRARKGSVPN WQSWAYLPRW SVIRRDIKRF  
 201 TGCKAAICLP MR\*

Further work revealed the complete nucleotide sequence <SEQ ID 383>:

```

5      1 ATGTTTATGA ACAAATTTTC CCAATCCGGA AAAGGTCTGT CCGGTTTTTT
      51 CTTCGGTTTG ATACTGGCGA CGGTCATTAT TGCCGGTATT TTGTTTATC
     101 TGAACCAGAG CGGTCAAAAT GCGTTCAAAA TCCCGGCTTC GTCGAGCAG
     151 CCTGCAGAAA CGGAAATCCT GAAACCGAAA AACCAGCCTA AGGAAGACAT
     201 CCAACCTGAA CCGGCCGATC AAAACGCCTT GTCCGAACCG GATGCTGCGA
10     251 CAGAGGCAGA GCAGTCGGAT GCGGAAAAAG CTGCCGACAA GCAGCCCGTT
     301 GCCGATAAAG CCGACGAGGT TGAAGAAAAG GCGGGCGAGC CGGAACGGGA
     351 AGAGCCGGAC GGACAGGCAG TGCCTAAGAA AGCGCTGACG GAAGAGCGTG
     401 AACAAACCGT CAGGGAAAAA GCGCAGAAGA AAGATGCCGA AACGGTTAAA
     451 AAACAAGCGG TAAACCGCT TAAAGAAACA GAGAAAAAAG CTTCAAAAGA
15     501 AGAGAAAAAG GCGGCGAAGG AAAAAGTTGC ACCCAAACCA ACCCGGAAC
     551 AAATCCTCAA CAGCGGCAGC ATCGAAAAAG CGCGCAGTGC CGCCGCCAAA
     601 GAAGTGCAGA AAATGAAAAC GTCCGACAAG GCGGAAGCAA CGCATTATCT
     651 GCAATGGGC GCGTATGCCG ACCGTCAGAG CGCGGAAGGG CAGCGTGCCA
     701 AACTGGCAAT CTTGGGCATA TCTTCCAAGG TGGTCGGTTA TCAGCGGGGA
20     751 CATAAACGCG TTTACCGGGT GCAAAGCGGC AATATGTCTG CCGATGCGGT
     801 GAAAAAATG CAGGACGAGT TGA AAAAACA TGAAGTCGCC AGCCTGATCC
     851 GTTCTATCGA AAGCAAATAA
  
```

This corresponds to the amino acid sequence <SEQ ID 384; ORF65-1>:

```

25     1 MFMNKFSQSG KGLSGFFFL ILATVIIAGI LFYLNQSGQN AFKIPASSKQ
      51 PAETEILKPK NQPKEDIQPE PADQNALSEP DAATEAEQSD AEKAADKQPV
     101 ADKADEVEEK AGEPEREEDP QAVRKKALT EEREQTVREK AQKKAETVK
     151 KQAVKPSKET EKKASKEEK AAKEKVAPKP TPEQILNSGS IEKARSAAAK
     201 EVQKMKTSK AEATHYLQMG AYADRQSAEG QRAKLAILGI SSKVVGQAG
     251 HKTLRYVQSG NMSADAVKKM QDELKHEVA SLIRSIK*
  
```

30 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF65 shows 92.0% identity over a 150aa overlap with an ORF (ORF65a) from strain A of *N. meningitidis*:

```

35      orf65.pep      10      20      30
                        ILKPHNQLKEDIQPD PADQNALSEPDAATE
                        |||:| | | | | | | | | | | | | | | | |
     orf65a      IIAGILFYLNQSGQNAFKIPVPSKQPAETEILKPKNQPKEDIQPEPADQNALSEPDAAKE
                        30      40      50      60      70      80

40      orf65.pep      40      50      60      70      80      90
                        AEQSDAENAADKQPVADKADEVEEKAGEPEREEDPGQAVRKKALTEEREQTVREKAQKKD
                        |||:| | | | | | | | | | | | | | | | | | | | | | | | | | |
     orf65a      AEQSDAENAADKQPVADKADEVEEKAGEPEREEDPGQAVRKKALTEEREQTVREKAQKKD
                        90      100     110     120     130     140

45      orf65.pep      100     110     120     130     140     150
                        AETVKIQAVKPSKETEEKASKEEKAAKEKVAPKPTPEQILNSGSIEXARSAAAKEVQKM
                        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
     orf65a      AETVKIQAVKPSKETEEKASKEEKAAKEKVAPKPTPEQILNSGSIEXARSAAAKEVQKM
                        150     160     170     180     190     200

50      orf65.pep      160     170     180     190     200     210
                        XNVRQGGXSR IICKWARMPTVRARKGSVPNWQSWAYLPRWSVIRRDIKRFTGCKAAICLP

55      orf65a      KTPDKAEATHYLQMGAYARRSAEQRAKLAILGISSKVVGQAGHKTLRYVQSGNMSAD
                        210     220     230     240     250     260
  
```

The complete length ORF65a nucleotide sequence <SEQ ID 385> is:

```

1 ATGTTTATGA ACAAATTTTC CCAATCCGGA AAAGGTCTGT CCGGTTTTTT
51 CTTCGGTTTG ATACTGGCGA CGGTCATTAT TGCCGGTATT TTGTTTATC
  
```

-244-

101 TGAACCAGAG CGGTCAAAAT GCGTTCAAAA TCCCGGTTCC GTCGAAGCAG  
 151 CCTGCAGAAA CGGAAATCCT GAAACCGAAA AACCAGCCTA AGGAAGACAT  
 201 CCAACCTGAA CCGGCCGATC AAAACGCCTT GTCCGAACCG GATGCTGCGA  
 5 251 AAGAGGCAGA GCAGTCGGAT GCGGAAAAAG CTGCCGACAA GCAGCCCCTT  
 301 GCCGACAAAAG CCGACGAGGT TGAGGAAAAG GCGGACGAGC CGGAGCGGGA  
 351 AAAGTCGGAC GGACAGGCAG TGGCAAGAA AGCACTGACG GAAGAGCGTG  
 401 AACAAACCGT CGGGGAAAAA GCGCAGAAGA AAGATGCCGA AACGGTTAAA  
 451 AAACAAGCGG TAAAACCATC TAAAGAAACA GAGAAAAAAG CTTCAAAAGA  
 501 AGAGAAAAAG GCGGAGAAGG AAAAAGTTGC ACCCAAACCG ACCCCGGAAC  
 10 551 AAATCCTCAA CAGCGGCAGC ATCGAAAAAG CGCGCAGTGC CGCTGCCAAA  
 601 GAAGTCGAGA AAATGAAAC GCCCGACAAG GCGGAAGCAA CGCATTATCT  
 651 GCAAATGGGC GCGTATGCCG ACCGCCGGAG CGCGGAAGGG CAGCGTGCCA  
 701 AACTGGCAAT CTTGGGCATA TCTTCCAAGG TGGTCGGTTA TCAGGCGGGA  
 751 CATAAACCGC TTTACCGGGT GCAAAGCGGC AATATGTCTG CCGATGCGGT  
 15 801 GAAAAAATG CAGGACGAGT TGAAAAACA TGAAGTCGCC AGCCTGATCC  
 851 GTTCTATCGA AAGCAAATAA

This encodes a protein having amino acid sequence <SEQ ID 386>:

1 MFMNKFSQSG KGLSGFFFL ILATVIIAGI LFYLNQSGQN AFKIPVPSKQ  
 51 PAETEILKPK NQPKEDIQPE PADQNALSEP DAAKEAEQSD AEKAADKQPV  
 20 101 ADKADEVEEK ADEPEREKSD QQAVRKKALT EEREQTVGEK AQKKDAETVK  
 151 KQAVKPSKET EKKASKEEKK AEKEKVAPKP TPEQILNSGS IEKARSAAK  
 201 EVQKMKT PDK AEATHYLQMG AYADRRSAEG QRAKLAILGI SSKVVG YQAG  
 251 HKTLYRVQSG NMSADAVKKM QDELKKHEVA SLIRSIESK\*

ORF65a and ORF65-1 show 96.5% identity in 289 aa overlap:

25		10	20	30	40	50	60
	orf65a.pep	MFMNKFSQSGKLSGFFFLILATVIIAGILFYLNQSGQNAFKIPVPSKQPAETEILKPK					
	orf65-1	MFMNKFSQSGKLSGFFFLILATVIIAGILFYLNQSGQNAFKIPASSKQPAETEILKPK					
30		10	20	30	40	50	60
	orf65a.pep	NQPKEDIQPEPADQNALSEPDAAKEAEQSDAEKAADKQPVADKADEVEEKADEPEREKSD					
	orf65-1	NQPKEDIQPEPADQNALSEPDAATEAEQSDAEKAADKQPVADKADEVEEKAGEPEREEDPD					
35		70	80	90	100	110	120
	orf65a.pep	NQPKEDIQPEPADQNALSEPDAAKEAEQSDAEKAADKQPVADKADEVEEKADEPEREKSD					
	orf65-1	NQPKEDIQPEPADQNALSEPDAATEAEQSDAEKAADKQPVADKADEVEEKAGEPEREEDPD					
40		130	140	150	160	170	180
	orf65a.pep	GQAVRKKALTEEREQTVGEKAQKKDAETVKKQAVKPSKETEKASKEEKKAEKEKVAPKP					
	orf65-1	GQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETEKASKEEKKAAKEKVAPKP					
45		190	200	210	220	230	240
	orf65a.pep	TPEQILNSGSIEKARSAAAKEVQKMKT PDKAEATHYLQMGAYADRRSAEQRAKLAILGI					
	orf65-1	TPEQILNSGSIEKARSAAAKEVQKMKTSDKAEATHYLQMGAYADRQSAEQRAKLAILGI					
50		250	260	270	280	290	
	orf65a.pep	SSKVVG YQAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX					
	orf65-1	SSKVVG YQAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX					
		250	260	270	280	290	

# 55 Homology with a predicted ORF from *N.gonorrhoeae*

ORF65 shows 89.6% identity over a 212aa overlap with a predicted ORF (ORF65.ng) from *N. gonorrhoeae*:

60		30	40	50	60	70	80
	ORF65ng	IIAGILLYLNQGGQNAFKIPAPSKQPAETEILKLNQPKEDIQPEPADQNALSEPDAVE					
	ORF65	ILKPHNQLKEDIQPD PADQNALSEPDAATE					
		10	20	30			

-245-

		90	100	110	120	130	140
	ORF65ng	AEQSDAEKAADKQPVADKADEVEEKAGEPEREEDGQAVRKKALTEEREQTVREKAQKKD					
		:     :     :     :     :     :     :     :     :     :     :     :					
5	ORF65	AEQSDAENAADKQPVADKADEVEEKAGEPEREEDGQAVRKKALTEEREQTVREKAQKKD					
		40	50	60	70	80	90
		150	160	170	180	190	200
	ORF65ng	AETVKKKAVKPSKETEKKASKEEKKAAKEKVAPKPTPEQILNSRSIEKARSAAAKEVQKM					
		:     :     :     :     :     :     :     :     :     :					
10	ORF65	AETVKIQAVKPSKETEKKASKEEKKAAKEKVAPKPTPEQILNSGSIEKARSAAAKEVQKM					
		100	110	120	130	140	150
		210	220	230	240	250	260
	ORF65ng	KNFGQGSQRIICKWARMNPNFGARKGSVPNWQSWAYLPKWSAIRRDIKRTACKAAICPP					
		:     :     :     :     :     :     :     :     :     :					
15	ORF65	XNVRQGSXRIICKWARMPTVRARKGSVPNWQSWAYLPRWSVIRRDIKRTGCKAAICLP					
		160	170	180	190	200	210
	ORF65ng	MR					
20	ORF65	MR					

An ORF65ng nucleotide sequence <SEQ ID 387> was predicted to encode a protein having amino acid sequence <SEQ ID 388>:

	1	MFMNKFSQSG	KGLSGFFFL	ILATVIIAGI	LLYLNQGGQN	AFKIPAPSKQ
25	51	PAETEILKLK	NQPKEDIQPE	PADQNALSEP	DVAKEAEQSD	AEKAADKQPV
	101	ADKADEVEEK	AGEPEREEDP	GQAVRKKALT	EEREQTVREK	AQKKDAETVK
	151	KKAVKPSKET	EKKASKEEKK	AAKEKVAPKP	TPEQILNSRS	IEKARSAAAK
	201	EVQKMNFGQ	GSQRIICKW	ARMPNPGARK	GSVPNWQSWA	YLPKWSAIRR
	251	DIKRTACKA	AICPPMR*			

30 After further analysis, the complete gonococcal DNA sequence <SEQ ID 389> was found to be:

	1	ATGTTTATGA	ACAAATTTTC	CCAATCCGGA	AAAGGTCTGT	CCGGTTTCTT
	51	CTTCGGTTTG	ATACTGGCAA	CGGTCATTAT	TGCCGGTATT	TTGCTTTATC
	101	TGAACAGGG	CGGTCAAAAT	GCGTCAAAA	TCCCGGCTCC	GTCGAAGCAG
35	151	CCTGCAGAAA	CGGAAATCCT	GAAACTGAAA	AACCAGCCTA	AGGAAGACAT
	201	CCAACCTGAA	CCGGCCGATC	AAAACGCCTT	GTCCGAACCG	GATGTTGCGA
	251	AAGAGGCAGA	GCAGTCGGAT	GCGGAAAAAG	CTGCCGACAA	GCAGCCCCTT
	301	GCCGACAAag	cgcacgAGGT	TGAAGAAAag	GcGGgcgAgc	cggAACGGga
	351	aGAGCCGGAC	ggACAGGCAG	TGCGCAAGAA	AGCACTGAcg	gAAGAgcGTG
	401	AACAAACcgt	cagggAAAAA	GCGCagaaga	AAGATGCCGA	AACGgTTAAA
40	451	AAacaaGCgg	tAaaaccgtc	tAAAGAAACa	gagaaaaaag	cTtcaaaaga
	501	agagaaaaag	gcggcgaaag	aaaAAGttgc	acccaaaccg	accccggaac
	551	aaatcctcaa	cagccgCagc	atcgaaaaag	cgcgtagtgc	cgctgccaaa
	601	gaAgtgcaGA	AAatgaaaaa	ctTtgggcaa	ggcgGaagcc	aacgcattat
	651	CTGcaaatgg	gcgcgtatgc	cgaccgtccg	gagcgcggaA	gggcagcgtg
45	701	ccaaACtggc	aAtcttgGgc	atatctTccg	aagtggctcgG	CTATCAGGCG
	751	GGACATAAAA	CGCTTTACCG	CGTGCAAagc	GGCAatatgt	ccgccgatgc
	801	gGTGAAAAAA	ATGCAGGACG	AGTTGAAAAA	GCATGGGGtt	gcCAGCCTGA
	851	TCCGTGcgAT	TGAAGGCAAA	TAA		

This encodes the following amino acid sequence <SEQ ID 390>:

	1	MFMNKFSQSG	KGLSGFFFL	ILATVIIAGI	LLYLNQGGQN	AFKIPAPSKQ
50	51	PAETEILKLK	NQPKEDIQPE	PADQNALSEP	DVAKEAEQSD	AEKAADKQPV
	101	ADKADEVEEK	AGEPEREEDP	GQAVRKKALT	EEREQTVREK	AQKKDAETVK
	151	KQAVKPSKET	EKKASKEEKK	AAKEKVAPKP	TPEQILNSRS	IEKARSAAAK
	201	EVQKMNFGQ	GSQRIICKW	ARMPTVRS	AEQRAKLAILG	ISSEVVGYYQA
55	251	GHKTLYRVQS	GNMSADAVKK	MQDELKKHGV	ASLIRAIEGK	*

ORF65ng-1 and ORF65-1 show 89.0% identity in 290 aa overlap:

		10	20	30	40	50	60
	orf65-1.pep	MFMNKFSQSGKGLSGFFFLILATVIIAGILFYLNQSGQNAFKIPASSKQPAETEILKPK					
		:     :     :     :     :     :     :     :     :     :					
60	orf65ng-1	MFMNKFSQSGKGLSGFFFLILATVIIAGILLYLNQGGQNAFKIPAPSKQPAETEILKLK					
		10	20	30	40	50	60

10

15

20

25 On this basis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

30 391>

35

40

45 This corresponds to the amino acid sequence <SEQ ID 392; ORF103>:

50

Further work elaborated the DNA sequence <SEQ ID 393> as:

55

-247-

5  
 351 GCGGAACCTG AACCCGATAC TCAACCGGCT GTTACCCATA AAATCCATAC  
 401 CCGCCTGCCT TCGGGTCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTG  
 451 GTTACAGCG CGTCGCTTTA CGCGCTGGGA AGCGGTAGTG CGGCAACGGG  
 501 CGGGTTATAT ATGCTTGCCCT TTGCACTGGG TACGCTGCCC AATCTTTTAG  
 551 CAATCGGCAT TTTTCCCTG CAACTGAAAA AAATCATGCA AAACCGATAT  
 601 ATCCGCCTGT GTACGGGATT ATCCGTATCA TTATGGGCAT TATGGAAACT  
 651 TGCCGTCCTG TGGCTGTAA

This corresponds to the amino acid sequence <SEQ ID 394; ORF103-1>:

10  
 1 MNHDITFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPI NRFWLILLN  
 51 TGRVSSYTAI GLILGLIGQV GVSLDQTRVL QNILYTAANL LLLFLGLYLS  
 101 GISSLAAKIE KIGKPIWRNL NPILNRLLPI KSIPACLA VG ILWGWLP CGL  
 151 VYSASLYALG SGSAATGGLY MLAFALGTLP NLLAIGIFSL QLKKIMQNR  
 201 IRLCTGLSVS LWALWKLAVL WL\*

Computer analysis of this amino acid sequence gave the following results:

15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF103 shows 93.8% identity over a 222aa overlap with an ORF (ORF103a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
20	orf103.pep	MNHDITFLTLFLLGXFGGTHCIGMCGGLSSAFXXQLPPHINRFWLILLNLTGRVSSYTAI					
	orf103a	MNXDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPIHNRXWLILLNLTGRVSSYTAI					
		10	20	30	40	50	60
25	orf103.pep	70	80	90	100	110	120
	orf103a	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL					
		70	80	90	100	110	120
30	orf103.pep	130	140	150	160	170	180
	orf103a	NPILNRLLPIKSIPACLA VGILWGWLP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP					
		130	140	150	160	170	180
35	orf103.pep	190	200	210	220		
	orf103a	NLLAIGIFSLQLXKIMQNRIRLCTGLSVSLWALWKLAVLWLX					
40		190	200	210	220		

The complete length ORF103a nucleotide sequence <SEQ ID 395> is:

45  
 1 ATGAACCANG ACATCACTTT CCTCACCTG TTCCTACTCG GTTTCTTCGG  
 51 CGGAACGCAC TGCATCGGTA TGTGCGGCGG ATTAAGCAGC GCGTTTGCGC  
 101 TCCAACCTCC CCGCATATC AACCGCTTNT GGCTGATCCT GCTGCTTAAC  
 151 ACAGGACGGG TAAGCAGCTA TACGGCAATC GGCCTGATAC TCGGATTAAT  
 201 CGGACAGGTC GCGGTTTCAC TCGACCAAAC CCGCGTCNTG CAGAATATTT  
 251 TATACACGGC CGCCAACCTC CTGCTGCTCT TTTTAGGCTT ATACTTGAGC  
 301 GGTATTTCTT CCTTGCGGCG AAAAATCGAG AAAATCGGCA AACCGATATG  
 351 GCGGAACCTG AACCCGATAC TCAACCGGCT GTTACCCATA AAATCCATAC  
 401 CCGCCTGCCT TCGGGTCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTA  
 451 GTTACAGCG CGTCGCTTTA CGCGCTGGGA AGCGGTAGTG CGGCAACGGG  
 501 CGGGTTATAT ATGCTTGCCCT TTGCACTGGG TACGCTGCCC AATCTTTNGG  
 551 CAATCGGCAT TTTTCCCTG CAACTGNAAA AAATCATGCA AAACCGATAT  
 601 ATCCGCCTGT GTACGGGATT ATCCGTATCA TTATGGGCAT TATGGAAACT  
 651 TGCCGTCCTG TGGCTGTAA

This encodes a protein having amino acid sequence <SEQ ID 396>:

1 MNXDITFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPI NRXWLILLN  
 51 TGRVSSYTAI GLILGLIGQV GVSLDQTRVX QNILYTAANL LLLFLGLYLS  
 101 GISSLAAKIE KIGKPIWRNL NPILNRLLPI KSIPACLA VG ILWGWLP CGL

151 VYSASLYALG SGSAATGGLY MLAFALGTLP NLXAIGIFSL QLXKIMQNR  
 201 IRLCTGLSVS LWALWKLAVL WL\*

ORF103a and ORF103-1 show 97.7% identity in 222 aa overlap:

5	orf103a.pep	10	20	30	40	50	60
		MNXDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRXWLILLNTGRVSSYTAI					
	orf103-1	10	20	30	40	50	60
		MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLILLNTGRVSSYTAI					
10	orf103a.pep	70	80	90	100	110	120
		GLILGLIGQVGVSLDQTRVXQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL					
	orf103-1	70	80	90	100	110	120
		GLILGLIGQVGVSLDQTRVLQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL					
15	orf103a.pep	130	140	150	160	170	180
		NPILNRLPIKSIPACLA VGILWGWLP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP					
	orf103-1	130	140	150	160	170	180
		NPILNRLPIKSIPACLA VGILWGWLP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP					
20	orf103a.pep	190	200	210	220		
		NLXAIGIFSLQLXKIMQNRIRLCTGLSVSLWALWKLAVLWLX					
	orf103-1	190	200	210	220		
		NLLAIGIFSLQLKKIMQNRIRLCTGLSVSLWALWKLAVLWLX					

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF103 shows 95.5% identity over a 222aa overlap with a predicted ORF (ORF103.ng) from *N.*

30 *gonorrhoeae*:

30	orf103.pep	MNHDITFLTLFLLGXFGGTHCIGMCGGLSSAFXXQLPPHINRFWLILLNTGRVSSYTAI	60
	orf103ng	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLILLNTGRISSTYTAI	60
35	orf103.pep	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL	120
	orf103ng	GLMLGLIGQLGISLDQTRVLQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL	120
40	orf103.pep	NPILNRLPIKSIPACLA VGILWGWLP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP	180
	orf103ng	NPILNRLPIKSIPACLA VGILWGWLP CGLVYSASLYALGSGSATTGGLYMLAFALGTLP	180
45	orf103.pep	NLLAIGIFSLQLXKIMQNRIRLCTGLSVSLWALWKLAVLWL	222
	orf103ng	NLLAIGIFSLQLKKIMQNRIRLCTGLSVSLWALWKLAVLWL	222

The complete length ORF103ng nucleotide sequence <SEQ ID 397> is:

1	ATGAACCACG	ACATCACTTT	CCTCACCCCTG	TTCCTGCTCG	GTTTCTTCGG
51	CGGAATCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GCGTTTGC
101	TCCAATCCC	CCCGCATATC	AACCGCTTTT	GGCTGATTCT	GCTGCTTAAC
151	ACAGGACGGA	TAAGCAGCTA	TACGGCAATC	GGCCTGATGC	TCGGATTAAT
201	CGGACAACTC	GGCATTTTCA	TCGACCAAAc	ccgcgTCCTG	CAAAATATTT
251	tatacacagc	ctccaaCCTC	CTGCTGCTCT	TTTTAGGCTT	ATACTTGAGC
301	GGTATTCTT	CCTTGCGGC	AAAAATCGAG	AAAATCGGCA	AACCGATATG
351	GCGCAACCTG	AACCGGATAC	TCAACCGGCT	GCTGCCATA	AAATCCATAC
401	CCGCCTGCCT	TGCTGTCCGA	ATATTATGGG	GCTGGCTGCC	GTGCGGACTG
451	GTTTACAGCG	CATCACTTTA	CGCGCTGGGA	AGCGGTAGTG	CGACAACCGG
501	CGGACTGTAT	ATGCTTGCTT	TTGCACTGGG	TACGCTGCCC	AATCTTTTGG
551	CAATCGGCAT	TTTTTCCCTG	CAACTGAAAA	AAATCATGCA	AAACCGATAT
601	ATCCGCCTGT	GTACAGGATT	ATCCGTATCA	TTATGGGCAT	TATGGAAGCT
651	TGCCGTCCTG	TGGCTGTAA			

This encodes a protein having amino acid sequence <SEQ ID 398>:

1 MNHDITFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPhi NRFWLILLLN  
 51 TGRISSTAI GLMLGLIGQL GISLDQTRVL QNILYTASNL LLLFLGLYLS  
 101 GISSLAAKIE KIGKPIWRNL NPILNRLPI KSIPACLA VG ILWGWLP CGL  
 151 VYSASLYALG SGSATTGGLY MLAFALGTLP NLLAIGIFSL QLKKIMQNR  
 201 IRLCTGLSVS LWALWKLAVL WL\*

In addition, ORF103ng and ORF103-1 show 97.3% identity in 222 aa overlap:

		10	20	30	40	50	60
10	orf103-1.pep	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPhINRFWLILLNTGRVSSYTAI					
	orf103ng	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPhINRFWLILLNTGRISSTAI					
		10	20	30	40	50	60
15	orf103-1.pep	70	80	90	100	110	120
	orf103ng	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL					
		70	80	90	100	110	120
20	orf103-1.pep	130	140	150	160	170	180
	orf103ng	NPILNRLPIKSIPACLA VGILWGWLP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP					
		130	140	150	160	170	180
25	orf103-1.pep	190	200	210	220		
	orf103ng	NLLAIGIFSLQLKKIMQNRIRLCTGLSVSLWALWKLAVLWLX					
		190	200	210	220		

30 Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 47

35 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 399>:

	1	ATGGA	AAACC	AAAGC	CGCT	CCTAG	GCTTT	CGCTT	GGCAC	TTTGG	CGGC
	51	GATGAC	GTGG	GGAAC	GCTGC	CGAT.	TCCGT	GCGGC	CAGTA	TTGAA	GTTTG
	101	TCGAT	GCGCC	GACGT	GCTG	TGGGT	GCGTT	TTACC	GTGGC	GGCGG	CGGTA
40	151	TTGTT	TGTTT	TGCTG	GCACT	GGGCG	GGCGG	CTGCC	GAAAGC	GGCGG	GGATT
	201	TTTCT	TGGTG	CTCAT	TCAAG	CTGCT	GCTGC	TCGGC	GTGGC	GGGCA	TTTTC
	251	GCAAA	CTTTG	TGCTG	ATTGC	CCAAG	GGCTG	CATTAT	ATTTT	CGCCG	ACCAC
	301	GACGC	CAGGT	TTGTG	GCGA	TTTCG	CCGTT	TACGAT	GATT	GTWGT	CGGTG
	351	TGTTG	GTGTT	TAAAG	ACCGG	ATGAC	TGCCG	CTCAG	AAAAT	CGGCT	TGGTT
45	401	TTGCT	GCTTG	CCGGT	TTTGT	TATGT	ATTTT	AACGA	TAAAT	TCGGC	GAGTT
	451	GTCGG	GTTTG	GGCGC	GTATG	C.AAG	GGCGT	GTTGT	GTGTG	GCGGC	AGGCA
	501	GTATG	GCGATG	GGTGT	GTAAT	GCCGT	GCGCG	AAAAG	CTGCT	GTCGG	CGCAA
	551	TTCGG	GCGCG	AACAG	ATTCT	GCTGT	TGATT	TATGC	GCGCA	GTGCC	CGCGT
	601	GTTCT	GCGCG	TTTGC	CGAAC	CGGC	CACAT	CGGA	AGTATG	GACGG	TACGT
50	651	TGGCG	TGGGT	ATGTAT	TGCG	TATTG	GCTGT	TGAAT	ACGTT	AATCG	GTATC
	701	GGCTC	GTTCG	GCGAG	GCGTT	GAAAC	ATTGG	GAGGC	TTCCA	AAGTC	AGCGC
	751	GGTAA	CAACC	TTGCT	CCCCG	TGTTT	ACCCT	AATAA	ATACT	TTGCT	CGGGC
	801	ATTAT	GTGAT	GCCTG	AAACT	TTTGC	CGCGC	CGGA.			

This corresponds to the amino acid sequence <SEQ ID 400; ORF104>:

	1	MENQR	PLLGF	RLALL	AAMTW	GTLPX	SVRQV	LKFVD	APTIV	WVRFT	VAAAV
55	51	LFVLL	ALGGR	LPKRR	DFSWC	SFRLL	LLLV	GISAN	FVLIA	QGLHY	ISPTT
	101	TQVLW	QISPF	TMIVV	GVLVF	KDRMT	AAQKI	GLVLL	LAGLL	MYFND	KFGEL
	151	SGLG	AYXKV	LLCA	AGSMA	VCNA	VAKLL	SAQFG	PQIL	LLIYA	ASAAV
	201	FLPFA	EPAHI	GSMD	GTLAW	CIAYC	CLNTL	IGYGS	FGEAL	KHWEA	SKVSA

251 VTTLLPVFTV INTLLGHYVM PETFAAP...

Further work revealed further partial DNA sequence <SEQ ID 401>:

```

1   ATGGAACACC AAAGCCGCT CCTAGGCTTC GCGTTGGCAC TTTTGGCGGC
5   51  GATGACGTGG GGAACGCTGC CGATTGCCGT GCGGCAGGTA TTGAAGTTTG
    101 TCGATGCGCC GACGCTGGTG TGGGTGCGTT TTACCGTGGC GCGGCGGGTA
    151 TTGTTTGTTC TGCTGGCACT GGGCGGGCGG CTGCCGAAGC GCGGGGATTG
    201 TTCTTGCTGC TCATTGAGGC TGCTGCTGCT CGGCGTGGCG GGCATTTTCGG
    251 CAACTTTTGT GCTGATTGCC CAAGGGCTGC ATTATATTTC GCCGACCACG
    301 ACGCAGGTTT TGTGGCAGAT TTCGCCGTTT ACGATGATTG TTGTCGGTGT
10   351 GTTGGTGTTT AAAGACCGGA TGACTGCCGC TCAGAAAATC GGCTTGGTTC
    401 TGCTGCTTGC CGGTTTGCTT ATGTTTTTTA ACGATAAATT CCGCGAGTTG
    451 TCGGGTTTGG GCGCGTATGC GAAGGGCGTG TTGCTGTGTG CCGCAGGCAG
    501 TATGGCATGG GTGTGTTATG CCGTGGCGCA AAAGCTGCTG TCGGCGCAAT
    551 TCGGGCCGCA ACAGATTCTG CTGTTGATTT ATGCGGCAAG TGCCGCCGTG
15   601 TTCCTGCCGT TTGCCGAACC GGCACACATC GGAAGTTTGG ACGGTACGTT
    651 GCGGTGGGTT TGTTTTTCGT ATTGCTGCTT GAATACGTTA ATCGGTTACG
    701 GCTCGTTCGG CGAGGCGTTG AAACATTGGG AGGCTTCCAA AGTCAGCGCG
    751 GTAACAACCT TGCTCCCCGT GTTACCCTGA ATAwTwWCTT TGCTCGGGCA
    801 TTATGTGATG CCTGAAACTT TTGCCGCGCC GGA...

```

20 This corresponds to the amino acid sequence <SEQ ID 402; ORF104-1>:

```

1   MENQRPLLGF ALALLAAMTW GTLPiAVRQV LKFVDAPTLV WVRFTVAAAV
5   51  LFVLLALGGR LPKRRDFSWC SFRLLLLGVA GISANFVLIA QGLHYISPPT
    101 TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLAGLL MFFNDKFGEL
    151 SGLGAYAKGV LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV
25   201 FLPPAEPAHI GSLDGT LAWV CFAYCCLNTL IGYGSFGEAL KHWEASKVSA
    251 VTTLLPVFTV IXXLLGHYVM PETFAAP...

```

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical HI0878 protein of *H. influenzae* (accession number U32769)

ORF104 and HI0878 show 40% aa identity in 277aa overlap:

```

30   orf104  4   QRPLLGFRLALLAAMTWGTLPSXSVRQVLKFVDAPTLVWXXXXXXXXXXXXXXXXXXXXXP- 62
        Q+PLLGF AL+ AM WG+LP +++QVL ++A T+VW P
    HI0878  3   QQPLLGFTFALITAMAWGSLPIALKQVLSVMNAQTIVWYRFIIAAVSLALLAYKKQLPE 62
35   orf104  63  --KRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPPTTQVLWQISPFMTIVVGVLVF 120
        K R ++W ++L+GV G+++NF+L + L+YI P+ Q+ +S F M++ GVL+F
    HI0878  63  LMKVRQYAW----IMLIGVIGLTSNFLFSSSLNYIEPSVAQIFIHLSSFGMLICGVLI 118
40   orf104  121 KDRMTAAQKIXXXXXXXXXXXMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL 180
        K+++ QKI ++FND+F +GL Y GV+L G++ WV +AQKL+
    HI0878  119 KEKLGLHQKIGLFLLLIGLGLFFNDRFADFAGLNQYSTGVILGVGGALIWVAYGMAQKLM 178
45   orf104  181 SAQFGPQQILLIYAASAAVFLPPAEPAHIGSMDGT LAWVCIAYCCLNTLIGYGSFGEAL 240
        +F QQILL++Y A F+P A+ + + + LA +C YCCLNTLIGYGS+ EAL
    HI0878  179 LRKFNSQQILLMMYLGCIAFMPMAFDSQVQELT-PLALICFIYCCLNTLIGYGSYAEAL 237
50   orf104  241 KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP 277
        W+ SKVS V TL+P+FT++ + + HY P FAAP
    HI0878  238 NRWDVSKVSVVITLVPLFTILFSLIAHYFSPADFAAP 274

```

50 Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF104 shows 95.3% identity over a 277aa overlap with an ORF (ORF104a) from strain A of *N. meningitidis*:

```

55   orf104.pep  10      20      30      40      50      60
        MENQRPLLGFRLALLAAMTWGTLPSXSVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR
        |||||  |||||  |||||  |||||  |||||  |||||
    orf104a    10      20      30      40      50      60
        MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR
        70      80      90     100     110     120

```



-251-

```

5  orf104.pep  LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF
   orf104a    LPKWRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF
   70          80          90          100         110         120

10  orf104.pep  KDRMTAAQKIGLVLLLAGLLMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL
   orf104a    KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL
   130         140         150         160         170         180

15  orf104.pep  SAQFGPQQIILLIYAASAAVFLPFAEPAHIGSMDGTLAWVCIAYCCLNTLIGYGSFGEAL
   orf104a    SAQFGPQQIILLIYAASAAVFLPFAELAHIGSLDGTALAWVCFAYCCLNTLIGYGSFGEAL
   190         200         210         220         230         240

20  orf104.pep  KHWEASKVSAVTTLPLPVFTVINTLLGHYVMPETFAAP
   orf104a    KHWEASKVSAVTTLPLPVFTVIFSLGHYVMPDTFAAPDMNGLGYAGALVVVGAVTAAVG
   250         260         270         280         290         300

```

The complete length ORF104a nucleotide sequence <SEQ ID 403> is:

```

25  1  ATGGAACACC AAAGCCGCT CCTAGGCTTC GCGTTGGCAC TTTTGGCGGC
   51  GATGACGTGG GGAACGCTGC CGATTGCCGT GCGGCAGGTA TTGAAGTTTG
  101  TCGATGCGCC GACGCTGGTG TGGGTGCGTT TTACCGTGGC GCGGCGGTA
  151  TTGTTGTGTT TGCTGGCATT GGGCGGGCGG CTGCCGAAGT GCGGGGATT
  201  TTCTTGGTGC TCATTCAAGC TGCTGCTGCT CGGCGTGGCG GGCATTTCGG
  251  CAACTTTGT GCTGATTGCC CAAGGGCTGC ATTATATTTC GCCGACCACG
  301  ACGCAGGTTT TGTGCGAGAT TTCGCCGTTT ACGATGATTG TTGTCGGTGT
  351  GTTGGTGTTT AAAGACCGGA TGACTGCCGC TCAGAAAATC GGCTTGGTGT
  401  TGCTGCTTGC CGGTTTGCTT ATGTTTTTTA ACGATAAATT CCGCGAGTTG
  451  TCGGGTTTGG GCGCGTATGC GAAGGGCGTG TTGCTGTGTG CCGCAGGCAG
  501  TATGGCATGG GTGTGTTATG CCGTGGCGCA AAAGCTGCTG TCGGCGCAAT
  551  TCGGCGCGCA ACAGATTCTG CTGTTGATTT ATGCGGCAAG TGCCGCGCTG
  601  TTCCTGCCGT TTGCCGAAC GGCACACATC GGAAGTTTGG ACGGTACGTT
  651  GGCGTGGGTT TGTTTTCGCT ATTGCTGCTT GAATACGTTA ATCGGTTACG
  701  GCTCGTTCGG CGAGGCGTTG AAACATTGGG AGGCTTCCAA AGTCAGCGCG
  751  GTAACAACCT TGCTCCCCGT GTTACCCTA ATATTTTCTT TGCTCGGGCA
  801  TTATGTGATG CTTGATACTT TTGCCGCGCC GGATATGAAC GGTTTGGGTT
  851  ATGCCGGCGC ACTGGTCGTG GTCGGGGGTG CCGTTACGGC GCGGTGGGG
  901  GACAGGCTGT TCAAACGCCG CTAG

```

This encodes a protein having amino acid sequence <SEQ ID 404>:

```

45  1  MENQRPLLGF ALALLAAMTW GTLPIAVRQV LKFVDAPTLV WVRFTVAAAV
   51  LFLVLLALGGR LPKWRDFSWC SFRLLLLLVA GISANFVLIA QGLHYISPTT
  101  TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLAGLL MFFNDKFGEL
  151  SGLGAYAKGV LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV
  201  FLPAELAH I GSLDGTALAW CFAYCCLNTL IGYGSFGEAL KHWEASKVSA
  251  VTTLLPVFTV IFSLLGHYVM PDTFAAPDMN GLGYAGALVV VGGAVTAAVG
  301  DRLFKRR*

```

ORF104a and ORF104-1 show 98.2% identity in 277 aa overlap:

```

55  orf104a.pep  MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFDAPTLVWVRFTVAAAVLFLVLLALGGR
   orf104-1     MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFDAPTLVWVRFTVAAAVLFLVLLALGGR
   10          20          30          40          50          60

60  orf104a.pep  LPKWRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF
   orf104-1     LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF
   70          80          90          100         110         120

65  orf104a.pep  KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL

```

ORF104 shows 93.9% identity over a 277aa overlap with a predicted ORF (ORF104.ng) from *N. gonorrhoeae*:

20	orf104.pep	MENQRPLLGFRLALLAAMTWGTLPXSVRQVLKFVDAPTLVWVRFTVAAAVLFLVLLALGGR       :	60
	orf104ng	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFLVLLALGGR	60
25	orf104.pep	LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTMIVVGVLVF      :	120
	orf104ng	LPKRRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPTTTQVLWQISPFMTMIVVGVLVF	120
30	orf104.pep	KDRMTAAQKIGLVLLLAGLLMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL    :	180
	orf104ng	KDRMTAAQKIGLVLLLVGLLMMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL	180
35	orf104.pep	SAQFGPQQILLIYAASA AVFLPFAEPAHIGSMDGTLAWVCIAYCCLNTLIGYGSFGEAL      :     :	240
	orf104ng	SAQFGPQQILLIYAASA AVFLXAEPAHIGSLDGTAWVCFVYCCLNTLIGYGSFGEAL	240
	orf104.pep	KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP      :	277
	orf104ng	KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMNGLGYVGALVVVGGA VTA AVG	300

The complete length ORF104ng nucleotide sequence <SEQ ID 405> is predicted to encode a protein having amino acid sequence <SEQ ID 406>:

45

1	<u>MENORPLLG</u>	<u>ALALLAAMTW</u>	GTLPIAVRQV	LKFVDAPTLV	WVRETVAAAV
51	<u>LFVLLALGGR</u>	<u>LPKRRDFSWH</u>	<u>SFRLLLLGVT</u>	<u>GISANFVLIA</u>	<u>QGLHYISPTT</u>
101	<u>TQVLWQISPF</u>	<u>MTIVVGLVLF</u>	<u>KDRMTAAQKI</u>	<u>GLVLLLVGLL</u>	<u>MFNDKFGEL</u>
151	<u>SGLGAYAKGV</u>	<u>LLCAAGSMAW</u>	<u>VCYAVAQKLL</u>	<u>SAQFGPQQIL</u>	<u>LLFYAASAAV</u>
201	<u>FLXAEPAHI</u>	<u>GSLDGTLAUV</u>	<u>CFVYCCNLTL</u>	<u>IGYGSFGEAL</u>	<u>KHWEASKVSA</u>
251	<u>VTTLPLVFTV</u>	<u>IFSLLGHYVM</u>	<u>PDTFAAPDMN</u>	<u>GLGYVGALVV</u>	<u>VGGAVTAAVG</u>
301	<u>DRPFKRR*</u>				

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 407>:

50	1	ATGGA <del>AA</del> ACC	AAAGGCCGCT	CCTAGGCTTC	GCGTTGGCAC	TTTTGGCGGC
	51	GATGACGTGG	GGGACGCTGC	CGATTGCCGT	GCGGCAGGTA	TTGAAGTTTG
	101	TCGATGCGCC	GACGCTGGTG	TGGGTGCGTT	TTACCGTGGC	GGCGGCGGTA
	151	TTGTTTGT	TGCTGGCATT	GGGCGGGCGG	CTGCCGAAGC	GGCGGGATTT
55	201	TTCTTGGCAT	TCATTCAGGC	TGCTGCTGCT	CGCGCTGACG	GGCGGATTCGG
	251	CAAAC <del>TT</del> TGT	GCTGATTGCC	CAAGGGCTGC	ATTATATTTT	GCCGACCACG
	301	ACGCAAGGTT	TGTGGCAGAT	TTCCGCCGTT	ACGATGATTG	TTGTCCGGCT
	351	GTTGGTGTTT	AAAGACCGGA	tgactGCGCG	GCAGAAATCT	GGTTTGGGTT
60	401	TGCTGCTgtT	CGGTTgtGTT	ATGTTTttta	ACGACAAATT	CGCGCAGTTG
	451	TCGGGTTTGG	GCGCGTATGC	GAAGGGCGTG	TTGCTGTGTG	CGGCAGGCAG
	501	TATGGCCTGG	GTGTGTTATG	CCGTGGCGCA	AAAGCTGCTG	TCGGCGCAAT
	551	TCGGGGCCGCA	ACAGATTCTG	CTGTTGATTT	ATGCGGcaag	tgccgccGCTG
	601	TTCCtgcggT	TTGccgaaCT	GGCACACATC	GGAAGTTTgg	aCGGTTACGtt
	651	GGCGCTGGGTT	TGTTTTGTGT	ATTGCTGCTT	GAATACGTTA	ATCGGTTACG

701 GCTCGTTCGG CGAGGCGTTG AAACATTGGG AGGCTTCCAA AGTCAGCGCG  
 751 GTAACAACCT TGCTCCCCGT GTTTACCGTA ATATTTTCTT TGCTCGGGCA  
 801 TTATGTGATG CCTGATACTT TTGCCGCGCC GGATATGAAC GGTTGGGGT  
 851 ATGTCGGCGC ACTGGTCGTG GTCGGGGGTG CGGTTACGGC GCGGTGGGG  
 901 GACAGGCCGT TCAAACGCCG CTAG

This corresponds to the amino acid sequence <SEQ ID 408; ORF104ng-1>:

1 MENQRPLLGF ALALLAAMTW GTLPIAVRQV LKFVDAPTLV WVRFTVAAAV  
 51 LFVLLALGGR LPKRRDFSWH SFRLLLLGVT GISANFVLIA QGLHYISPTT  
 101 TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLVGLL MFFNDKFGEL  
 151 SGLGAYAKGV LLCAAGSMAW VCAYAQAQKLL SAQFGPQQIL LLIYAASAAV  
 201 FLPFAEPAHI GSLDGTAWV CFVYCCLNTL IGYGSFGEAL KHWEASKVSA  
 251 VTTLLPVFTV IFSLLGHYVM PDTFAAPDMN GLGYVGALVV VGGAVTAAVG  
 301 DRPFKRR\*

ORF104ng-1 and ORF104-1 show 97.5% identity in 277 aa overlap:

15	orf104-1.pep	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR
	orf104ng-1	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR
20	orf104-1.pep	LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF
	orf104ng-1	LPKRRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF
25	orf104-1.pep	KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCAYAQAQKLL
	orf104ng-1	KDRMTAAQKIGLVLLLVGLLMMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCAYAQAQKLL
30	orf104-1.pep	SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGTAWVCFAYCCLNTLIGYGSFGEAL
	orf104ng-1	SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGTAWVCFVYCCLNTLIGYGSFGEAL
35	orf104-1.pep	KHWEASKVSAVTTLLPVFTVIXXLLGHYVMPETFAAP
	orf104ng-1	KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMNGLGYVGALVVVGGAVTAAVG

In addition, ORF104ng-1 shows significant homology with a hypothetical *H. influenzae* protein:

gi|1573895 (U32769) hypothetical [Haemophilus influenzae] Length = 306  
 Score = 237 bits (598), Expect = 8e-62  
 Identities = 114/280 (40%), Positives = 168/280 (59%), Gaps = 8/280 (2%)

Query: 30 QRPXXXXXXXXXXMTWGTLPPIAVRQVLKFVDAPTLVWXXXXXXXXXXXXXXXXXP- 88  
 Q+P M WG+LPIA++QVL ++A T+VW P

Sbjct: 3 QQPLLGFTFALITAMAWGSLPIALKQVLSVMNAQTIVWYRFIIAAVSLALLAYKKQLPE 62

Query: 89 --KRRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF 146  
 K R ++W ++L+GV G+++NF+L + L+YI P+ Q+ +S F M++ GVL+F

Sbjct: 63 LMKVRQYAW----IMLIGVIGLTSNFLFSSSLNIEPSVAQIFIHLSSFGMLICGVLI 118

Query: 147 KDRMTAAQKIXXXXXXXXXMMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCAYAQAQKLL 206  
 K+++ QKI +FFND+F +GL Y+ GV+L G++ WV Y +AQKL+

Sbjct: 119 KEKLGHLHQKIGLFLLLIGLGLFFNDRDFAFAGLNQYSTGVILGVGALIWVAYGMAQKLM 178

Query: 207 SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGTAWVCFVYCCLNTLIGYGSFGEAL 266  
 +F QQILL++Y A F+P A+ + + L LA +CF+YCCLNTLIGYGS+ EAL

Sbjct: 179 LRKFNSQQILLMMYLGCAIAFMPMADFSQVQELT-PLALICFIYCCLNTLIGYGSYAEAL 237

Query: 267 KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMN 306

-252-

5	orf104-1	 KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL	130	140	150	160	170	180
	orf104a.pep	SAQFGPQQIILLIYAASA AVFLPFAELAHIGSLDGT LAWVCFAYCCLNTLIGYGSFGEAL	190	200	210	220	230	240
10	orf104-1	 SAQFGPQQIILLIYAASA AVFLPFAELAHIGSLDGT LAWVCFAYCCLNTLIGYGSFGEAL	190	200	210	220	230	240
	orf104a.pep	 KHWEASKVSAVTTLLPVFTVIFSL LGHYVMPDTFAAPDMNGLGYAGALVVVGAVTA AVG	250	260	270	280	290	300
15	orf104-1	 KHWEASKVSAVTTLLPVFTVIXXLLGHYVMPETFAAP	250	260	270			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF104 shows 93.9% identity over a 277aa overlap with a predicted ORF (ORF104.ng) from *N. gonorrhoeae*:

20	orf104.pep	MENQRPLLGFRLALLAAMTWGTLPXSVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR	60
	orf104ng	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR	60
25	orf104.pep	LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF	120
	orf104ng	LPKRRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF	120
30	orf104.pep	KDRMTAAQKIGLVLLLAGLLMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL	180
	orf104ng	KDRMTAAQKIGLVLLVGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL	180
35	orf104.pep	SAQFGPQQIILLIYAASA AVFLPFAEPAHIGSMDGT LAWVCIAYCCLNTLIGYGSFGEAL	240
	orf104ng	SAQFGPQQIILLIYAASA AVFLLXAEPAHIGSLDGT LAWVCFVYCCCLNTLIGYGSFGEAL	240
	orf104.pep	KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP	277
	orf104ng	KHWEASKVSAVTTLLPVFTVIFSL LGHYVMPDTFAAPDMNGLGYV GALVVVGAVTA AVG	300

The complete length ORF104ng nucleotide sequence <SEQ ID 405> is predicted to encode a protein having amino acid sequence <SEQ ID 406>:

	1	MENORPLLGF	ALALLAAMTW	GTLPIAVRQV	LKFVDAPTLV	WVRFTVAAAV
	51	LFVLLALGGR	LPKRRDFSWH	SFRLLLLGVT	GISANFVLIA	QGLHYISPTT
	101	TQVLWQISPF	TMIVVGVLVF	KDRMTAAQKI	GLVLLLVGLL	MFNDKFGEL
	151	SLGAYAKGV	LLCAAGSMAW	VCYAVAQKLL	SAQFGPQQIL	LLIYAASAAV
45	201	FLLXAEPAHI	GSLDGT LAWV	CFVYCCCLNTL	IGYGSFGEAL	KHWEASKVSA
	251	VTLLPVFTV	IFSL LGHYVM	PDTFAAPDMN	GLGYV GALVV	VGGAVTA AVG
	301	DRPFKRR*				

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 407>:

50	1	ATGGA AACC	AAAGGCCGCT	CCTAGGCTTC	GCGTTGGCAC	TTTGGCGGC
	51	GATGACGTGG	GGGACGCTGC	CGATTGCCGT	GCGGCAGGTA	TTGAAGTTTG
	101	TCGATGCGCC	GACGCTGGTG	TGGGTGCGTT	TTACCGTGGC	GGCGGCGGTA
	151	TTGTTTGT	TGCTGGCATT	GGGCGGCGG	CTGCCGAAGC	GGCGGGATTT
	201	TTCTTGGCAT	TCATT CAGGC	TGCTGCTGCT	CGGCGTGACG	GGCATTTCGG
55	251	CAAACTTTGT	GCTGATTGCC	CAAGGGCTGC	ATTATATTTC	GCCGACCACG
	301	ACGCAGGTTT	TGTGGCAGAT	TTCCGCCGTTT	ACGATGATTG	TTGTCGGCGT
	351	GTTGGTGT	AAAGACCGGA	tgaCTGCCGC	GCAGAAAATC	GGTTTGGTTT
	401	TGCTGCTtgT	CGGTTtgCTT	ATGTTTtTta	ACGACAAATT	CGGCAGATTG
	451	TCGGGTTTGG	GCGCGTATGC	GAAGGGCGTG	TTGCTGTGTG	CGGCAGGCAG
	501	TATGGCCTGG	GCTGTATTATG	CCGTGGCGCA	AAAGCTGCTG	TCCGCGCAAT
60	551	TCGGGCGGCA	ACAGATTCTG	CTGTTGATTT	ATGCGGcaag	tgccgccGTG
	601	TTCCtgccgT	TTGccgaacc	GGCACACATC	GGAAGTTTgg	aCGGTACGtt
	651	GCGGTGGGTT	TGTTTTGTGT	ATTGCTGCTT	GAATACGTTA	ATCGGTTACG

-253-

701 GCTCGTTCGG CGAGGCGTTG AACATTGGG AGGCTTCCAA AGTCAGCGCG  
 751 GTAACAACCT TGCTCCCCGT GTTTACCGTA ATATTTTCTT TGCTCGGGCA  
 801 TTATGTGATG CCTGATACTT TTGCCGCGCC GGATATGAAC GGTTGGGTT  
 851 ATGTCGGCGC ACTGGTCGTG GTCGGGGGTG CGGTTACGGC GGCGGTGGGG  
 901 GACAGGCCGT TCAAACGCCG CTAG

This corresponds to the amino acid sequence <SEQ ID 408; ORF104ng-1>:

1 MENQRPLLGF ALALLAAMTW GTLPIAVRQV LKFVDAPTLV WVRFTVAAAV  
 51 LFVLLALGGR LPKRRDFSWH SFRLLLLGVT GISANFVLIA QGLHYISPTT  
 101 TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLVGLL MFENDKFGEL  
 151 SGLGAYAKGV LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV  
 201 FLPFAEPAHI GSLDGTAWV CFVYCCLNTL IGYGSFGEAL KHWEASKVSA  
 251 VTTLLPVFTV IFSLLGHYVM PDTFAAPDMN GLGYVGALVV VGGAVTAAVG  
 301 DRPFKRR\*

ORF104ng-1 and ORF104-1 show 97.5% identity in 277 aa overlap:

15	orf104-1.pep	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR
	orf104ng-1	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR
20	orf104-1.pep	LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF
	orf104ng-1	LPKRRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF
25	orf104-1.pep	KDRMTAAQKIGLVLLLAGLIMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL
	orf104ng-1	KDRMTAAQKIGLVLLLVGLIMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL
30	orf104-1.pep	SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGTAWVCFAYCCLNTLIGYGSFGEAL
	orf104ng-1	SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGTAWVCFVYCCLNTLIGYGSFGEAL
35	orf104-1.pep	KHWEASKVSAVTTLLPVFTVIXXLLGHYVMPETFAAP
	orf104ng-1	KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMNGLGYVGALVVVGGAVTAAVG

In addition, ORF104ng-1 shows significant homology with a hypothetical *H.influenzae* protein:

gi|1573895 (U32769) hypothetical [Haemophilus influenzae] Length = 306  
 Score = 237 bits (598), Expect = 8e-62  
 Identities = 114/280 (40%), Positives = 168/280 (59%), Gaps = 8/280 (2%)

Query: 30 QRPXXXXXXXXXXMTWGTLPPIAVRQVLKFVDAPTLVWXXXXXXXXXXXXXXXXXXXXP- 88  
 Q+P M WG+LPIA++QVL ++A T+VW P

Sbjct: 3 QQPLLGFTFALITAMAWGSLPIALKQVLSVMNAQTIVWYRFIIAAVSLALLAYKKQLPE 62

Query: 89 --KRRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF 146  
 K R ++W ++L+GV G+++NF+L + L+YI P+ Q+ +S F M++ GVL+F

Sbjct: 63 LMKVRQYAW----IMLIGVIGLTSNFLFSSSLNYIEPSVAQIFIHLSSFGMLICGVLI 118

Query: 147 KDRMTAAQKIXXXXXXXXXXXMFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL 206  
 K+++ QKI ++FND+F +GL Y+ GV+L G++ WV Y +AQKL+

Sbjct: 119 KEKLGHLHQKIGLFLLLIGLGLFFNDRDFAAGLNQYSTGVILGVGGALIWVAYGMAQKLM 178

Query: 207 SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGTAWVCFVYCCLNTLIGYGSFGEAL 266  
 +F QQILL++Y A F+P A+ + + L LA +CF+YCCLNTLIGYGS+ EAL

Sbjct: 179 LRKFNSQQILLMYLGCAIAFMPMADFSQVQELT-PLALICFIYCCLNTLIGYGSYAEAL 237

Query: 267 KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMN 306

W+ SKVS V TL+P+FT++FS + HY P FAAP++N  
 Sbjct: 238 NRWDVSKVSVVITLVPLFTILFSHIAHYFSPADFAAPELN 277

Based on this analysis, including the presence of a putative leader sequence and several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from  
 5 *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 48

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 409>:

```

10      1  ATGGTAGCTC  GTCGGGCTCA  TAACCCGAAG  GTCGTAGGTT  CGAATCCTGT
      51  .CCCGCAACC  TAATTTCAAA  CCCCTCGGTT  CAATGCCGAG  GG.GTTTTGT
     101  T.TTGCCTGT  TTCCTGTTTC  CTGTTTCCTG  CCGCCTCCGT  TTTTGGCCGG
     151  ATTTTCCTTC  CGGCCGCAAT  ATCGGAACGG  CAGACCGCCG  TCTGTTTGCG
     201  GTTGCAAATT  CAGGCAGTTT  GGCTACAATC  TTCCGCATTG  TCTTCAAGAA
     251  AGCCAACCAT  GCCGACCGTC  CGTTTACCG  AATCCGTCAG  CAAACAAGAC
     15  301  CTTGATGCTC  TGTTCGAGTG  GGCAAAAGCA  AGTTACGGTG  CAGAAAGTTG
     351  CTGGAACACG  CTGTATCTGA  ACGGTCysCC  TTTGGGCAAC  CTGTCGCCGG
     401  AATGGGTGGA  ACGCGTsmmA  AAAGACTGGG  AGGCAGGCTG  CyCGGAGTCT
     451  TCAGACGCGA  TTTTCTGAA  TgCGGACGGc  TGgCctGATA  TGGgCGGAcg
     501  cTTACAGCAC  CTCGCCCTCG  GTTGGCACTG  TCGGGGGCTG  TTGGACGGst
     20  551  GCGCAACGA  GTGTTTCGAC  CTGACCGACG  GCGGCGGCAA  CCCCTTGTTC
     601  ACGCTCGaAc  GCGCCGyTTT  mCGTCTCTkTC  GGACTGCTCA  GCCGCGCCGT
     651  CCATCTCAAC  GGTCTGACCG  AATCGGACGG  CCGATGGCAT  TTCTGGATAG
     701  GCAGGCGCAG  TCCGCACAAA  GCAGTCGATC  CCAACAAACT  CGACAATACT
     751  rCCGCCGGCG  GTGTTCCGG  CGGCGAAATG  CCGTCTGAAG  CCGTGTGTCTG
     25  801  CGAAAGCAGC  GAAGAAGCCG  GTTGGATAA  AACGTGcTT  CCGCTCATCC
     851  GCCCGGTATC  GCAGCTGCAC  AGCCTGCGCT  CCGTCAGCCG  GGGTGTACAC
     901  AATGAAATCC  TGTATGTATT  CGATGCCGTC  CTGCCG...
```

This corresponds to the amino acid sequence <SEQ ID 410; ORF105>:

```

30      1  MVARRAHNPK  VVGSNPXPAT  XFQTPRFNAE  XVLXLPVSCF  LFPAASVFRCR
      51  IFLPAAISER  QTAVCLRLQI  QAVWLQSSAL  SSRKPTMPTV  RFTESVSKQD
     101  LDALFEWAKA  SYGAESCWKT  LYLNGXPLGN  LSPewVERVX  KDWEAGCXES
     151  SDGIFLNADG  WPDmGGRlQH  LALGWHCAGL  LDGWRNECFD  LTDGGGNPLF
     201  TLERAXXRPX  GLLSRaVHLN  GLTESDGRWH  FWIGRRSPHK  AVDPNKLDNT
     251  XAGGVSGGEM  PSEAVCRESS  EEAGLDKTL  PLIRPVSQLH  SLRSVSRGVH
     35  301  NEILYVFDAV  LP...
```

Further work revealed the complete nucleotide sequence <SEQ ID 411>:

```

40      1  ATGCCGACCG  TCCGTTTTAC  CGAATCCGTC  AGCAAACAAG  ACCTTGATGC
      51  TCTGTTCGAG  TGGGCAAAAG  CAAGTTACGG  TGCAGAAAGT  TGCTGGAAAA
     101  CGCTGTATCT  GAACGGTCTG  CCTTTGGGCA  ACCTGTCGCC  GGAATGGGTG
     151  GAACGCGTCA  AAAAAGACTG  GGAGGCAGGC  TGCTCGGAGT  CTTCAGACGG
     201  CATTTTCTG  AATGCGGACG  GCTGGCCTGA  TATGGGCGGA  CGCTTACAGC
     251  ACCTCGCCCT  CGGTTGGCAC  TGTGCGGGGC  TGTGGACGG  CTGGCGCAAC
     301  AGTGTTTCG  ACCTGACCGA  CGGCGGCGGC  AACCCCTTGT  TCACGCTCGA
     351  ACGCGCCGCT  TTCCGTCCTT  TCGGACTGCT  CAGCCGCGCC  GTCCATCTCA
     45  401  ACGGTCTGAC  CGAATCGGAC  GGCGATGGC  ATTTCTGGAT  AGGCAGGCGC
     451  AGTCCGCACA  AAGCAGTCGA  TCCCAACAAA  CTCGACAATA  CTGCCGCCGG
     501  CGGTGTTTCC  GCGGCGAA  TGCCGTCTGA  AGCCGTGTGT  CGCGAAAGCA
     551  CCGAAGAAGC  CGGTTGGAT  AAAACGCTGC  TTCCGCTCAT  CCGCCCGGTA
     601  TCGCAGCTGC  ACAGCTGCG  CTCCGTCAGC  CGGGGTGTAC  ACAATGAAAT
     50  651  CCTGTATGTA  TTCGATGCCG  TCCTGCCCGA  AACCTTCTG  CCTGAAATC
     701  AGGATGGCGA  AGTGGCGGGT  TTTGAGAAAA  TGGACATCGG  CCGTCTGTTG
     751  GATGCCATGT  TGTGCGGAAA  CATGATGCAC  GACGCGCAAC  TGGTTACGCT
     801  GGACGCGTTT  TGCCGTTACG  GTCTGATTGA  TGCCGCCCAT  CCGCTGTCCG
     851  AGTGGCTGGA  CGGCATACGT  TTATAG
```

55 This corresponds to the amino acid sequence <SEQ ID 412; ORF105-1>:

5

1	MPTVRFTESV	SKQDLDALE	WAKASYGAES	CWKTLYLNGL	PLGNLSPEWV
51	ERVKKDWEAG	CSESSDGIFL	NADGWPDMGG	RLQHLALGWH	CAGLLDGWRN
101	ECFDLTDGGG	NPLFTLRAA	FRPFGLLSRA	VHLNGLTESD	GRWFHWIGRR
151	SPHKAVDPNK	LDNTAAGGVS	GGEMPSEAVC	RESSEEAGLD	KTLLPLIRPV
201	SQHLSLRSVS	RGVHNEILYV	FDAPVLPETFL	PENQDGEVAG	EKKMDIGGLL
251	DAMLSGNNMMH	DAQVLTDAF	CRYGLIDAAH	PLSEWLDGIR	L*

**Homology with a predicted ORF from *N.meningitidis* (strain A)**

10 *meningitidis:*

40 The complete length ORF105a nucleotide sequence <SEQ ID 413> is:

This encodes a protein having amino acid sequence <SEQ ID 414>:

60 1 MPTVRFTESV SKHDLDALE WAKASYGAES CWKTLYLNGL PLGNLSPEWA  
51 ERVKKDWEAG CSESSDGIFL NADGWPDMDGR RLOHLARIWK EAGLLHGWRD

-256-

101 ECFDLTDGGS NPLFALERAA FRPFGLLSRA VHLNGLVESD GRWHFWIGRR  
 151 SPHKAVDPDK LDNTAAGGVS SGELPSETVC RESSEEAGLD KTLPLIRPV  
 201 SQLHSLRPVS RGVHNEILYV FDAVLPETFL PENQDGEVAG FEKMDIGGLL  
 251 AAMLSGNMMH DAQLVTLDAF CRYGLIDAAH PLSEWLDGIR L\*

5 ORF105a and ORF105-1 show 93.8% identity in 291 aa overlap:

		10	20	30	40	50	60
	orf105a.pep	MPTVRFTESVSKHDL	DALFEWAKASYGAESC	WKTLYLNGPLPLGNLS	PEWAERVVKDWEAG		
	orf105-1	MPTVRFTESVSKHDL	DALFEWAKASYGAESC	WKTLYLNGPLPLGNLS	PEWVERVVKDWEAG		
10		10	20	30	40	50	60
	orf105a.pep	CSESSDGIFLNADGW	PDMGRRQLQHLARI	WKEAGLLHGW	RDECFDLTDGGSN	NPLFALERAA	
	orf105-1	CSESSDGIFLNADGW	PDMGRRQLQHLALG	WHCAGLLDGWRNE	CFDLTDGGGNPL	FTLERAA	
15		70	80	90	100	110	120
	orf105a.pep	FRPFGLLSRAVHLN	GLVESDGRWHFWIG	RSPHKAVDPDKLD	NTAAGGVSSGELP	SETVC	
	orf105-1	FRPFGLLSRAVHLN	GLTESDGRWHFWIG	RSPHKAVDPNKL	DNTAAGGVSSGEM	PSEAVC	
20		130	140	150	160	170	180
	orf105a.pep	RESSEEAGLDKTL	PLIRPVSQLHSLRP	SVSRGVHNEILYV	FDVLPETFLPENQ	DGEVAG	
	orf105-1	RESSEEAGLDKTL	PLIRPVSQLHSLR	SVSRGVHNEILYV	FDVLPETFLPENQ	DGEVAG	
25		190	200	210	220	230	240
	orf105a.pep	FEKMDIGGLLAAM	LSGNMMHDAQLVT	LDAFCRYGLIDAA	HPLSEWLDGIRL	X	
	orf105-1	FEKMDIGGLLDAM	LSGNMMHDAQLVT	LDAFCRYGLIDAA	HPLSEWLDGIRL	X	
30		250	260	270	280	290	
	orf105a.pep						
	orf105-1						
35		250	260	270	280	290	

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF105 shows 87.5% identity over a 312aa overlap with a predicted ORF (ORF105.ng) from *N. gonorrhoeae*:

40	orf105.pep	MVARRAHNPKVVGSNPXPATXFQTPRFNAEXVLXLPVSCFLFPAASVFCRIFLPAAISER	60
	orf105ng	MVARRAHNPKVVGSNPAPATKYQTPRFNAEGVLF-----FLFPAASVFCRIFLPAAISER	55
45	orf105.pep	QTAVCLRLQIQAVWLQSSALSSRKPTMPTVRFTESVSKQDLDALEWAKASYGAESCWKT	120
	orf105ng	QAAVCLRLQIQAVWLQSSALCSRKPAMPTVRFTESVSKQDLDALEWAKASYGAESCWKT	115
50	orf105.pep	LYLNGXPLGNLSPEWVERVXKDWEAGCXESSDGIFLNADGWPDMMGRRQLQHLALGWHCAGL	180
	orf105ng	LYLNRLPLGNLSPEWAERIKKDWEAGCSESSNGIFLNADGWPDMMGRRQLQHLARTWNKAGL	175
55	orf105.pep	LDGWRNECFDLTDGGGNPLFTLERAXRXPGLLSRAVHLNGLTESDGRWHFWIGRRSPHK	240
	orf105ng	LHGWRNECFDLTDGGGNPLFTLERAAFRPFGLLIRAVHLNGLVESNGRWHFWIGRRSPHK	235
60	orf105.pep	AVDPNKLNTXAGGVSGGEMPSEAVCRESEEAGLDKTLPLIRPVSQLHSLRSVSRGVH	300
	orf105ng	AVDPGKLDNIAGGVSGGEMPSEAVCRESEEAGLDKTLPLIRPVSRHSLRPVSRGVH	295
	orf105.pep	NEILYVFDVLP	312
	orf105ng	NEILYVFDVLPETFLPENQDGEVAGFEKMDIGGLLDAMLSKNMMHDAQLVTLDAFYRYG	355

A complete length ORF105ng nucleotide sequence <SEQ ID 415> was predicted to encode a protein having amino acid sequence <SEQ ID 416>:



-257-

1 MVARRAHNPK VVGSNPAPAT KYQTPRFNAE GVLFFLPAA SVFCRIFLPA  
 51 AISERQAAVC LRLQIQAVWL QSSALCSRKP AMPTVRFTES VSKQDLDALE  
 101 ERAKASYGAE SCWKTLYLNR LPLGNLSPEW AERIKKDWEA GCSESSNGIF  
 151 LNADGWPDGM GRLQHLARTW NKAGLLHGWR NECFDLTDGG GNPLFTLERA  
 5 AFRPFGLLIR AVHLNGLVES NGRWHFWIGR RSPHKAVDPG KLDNIAGGGV  
 251 SGGEMPSEAV CRESSEAGL DKTFLPLIRP VSRHLSLRPV SRGVHNEILY  
 301 VFDAVLPETF LPENQDGEVA GFKMDIGGL LDAMLSKNMM HDAQLVTLDA  
 351 FYRYGLIDAA HPLSEWLDGI RL\*

Further work revealed the complete nucleotide sequence <SEQ ID 417>:

10 1 ATGCCGACCG TCCGTTTTAC CGAATCCGTC AGCAAACAAG ACCTTGATGC  
 51 CCTGTTTCGAG CGGGCAAAAG CAAGTTACGG TGCCGAAAGT TGCTGGAAAA  
 101 CGCTGTATCT GAACCGCTT CCTTTGGGCA ATCTGTCGCC GGAATGGGCT  
 151 GAGCGCATCA AAAAGACTG GGAGGCAGGC TGCTCCGAGT CTTAGACGG  
 201 CATTTTCTG AATGCGGACG GCTGGCCGA TATGGGCGGA CGCTTGACG  
 15 ACCTCGCCCG CACATGGAAC AAGGCGGGC TGCTTACGG ATGGCGCAAC  
 301 GAGTGTTCG ACCTGACCGA CGGCGGCGGC AACCCCTTGT TCACGCTCGA  
 351 ACGCGCCGCT TTCCGTCCTG TCGGACTACT CAGCCGCGCC GTCCATCTCA  
 401 ACGGTTTGGT CGAATCGAAC GGCAGATGGC ATTTTGGAT AGGCAGGCGC  
 451 AGTCCGCACA AAGCAGTCGa tCCGCGCAAG CTCGACAATA TTGCCGCGCG  
 20 CGGTGTTTCC GCGGCGGAAA TGCCGTCTGA AGCCGTGTGC CGCGAAAGCA  
 551 GCGAAGAAGC CGGTTTGGAT AAAACGCTGT TTCCGCTCAT CCGCCCAGTA  
 601 TCGCGGCTGC ACAGCCTTCG CCCCCTCAGC CGAGGTGTGC ACAATGAAAT  
 651 CCTGTATGTG TTCGATGCCG TCCTGCCCGA AACCTTCTG CCTGAAAATC  
 701 AGGATGGCGA GGTAGCGGGT TTTGAAAAGA TGGACATTGG CGGCCTATTG  
 25 GATGCCATGT TGTCGAAAAA CATGATGCAC GACGCGCAAC TGGTTACGCT  
 801 GGACGCGTTT TACCGTTACG GTCTGATTGA TGCCGCCCAT CCGCTGTCCG  
 851 AGTGGCTGGA CGGCATACGT TTATAG

This corresponds to the amino acid sequence <SEQ ID 418; ORF105ng-1>:

30 1 MPTVRFTESV SKQDLDALE RAKASYGAES CWKTLYLNR LPLGNLSPEWA  
 51 ERIKKDWEAG CSESSDGIFL NADGWPDMMG RLQHLARTWN KAGLLHGWRN  
 101 ECFDLTDGGG NPLFTLERA FRPFGLLSRA VHLNGLVESN GRWHFWIGRR  
 151 SPHKAVDPGK LDNIAGGGVS GGEMPSEAVC RESSEEAGLD KTLFPLIRPV  
 201 SRLHSLRPVS RGVHNEILYV FDAVLPETF PENQDGEVAG FEKMDIGGLL  
 251 DAMLSKNMMH DAQLVTLDAF YRYGLIDAAH PLSEWLDGIR L\*

35 ORG105ng-1 and ORF105-1 show 93.5% identity in 291 aa overlap:

	10	20	30	40	50	60
orf105-1.pep	MPTVRFTESVSKQDLDALE	FEWAKASYGAES	SCWKTLYLNL	LPLGNLSPEW	VERVKKDWEAG	
orf105ng-1	MPTVRFTESVSKQDLDALE	FERAKASYGAES	SCWKTLYLNR	LPLGNLSPEW	AERIKKDWEAG	
	10	20	30	40	50	60
orf105-1.pep	CSESSDGIFLNADGWPDMMG	RQLQHLALGWH	CAGLLDGRNE	CFDLTDGGGN	NPLFTLERA	
orf105ng-1	CSESSDGIFLNADGWPDMMG	RQLQHLARTWN	KAGLLHGWRNE	CFDLTDGGGN	NPLFTLERA	
	70	80	90	100	110	120
orf105-1.pep	FRPFGLLSRAVHLNGLTES	DGRWHFWIGRR	SPHKAVDPN	KLDNTAAGV	SGGEMPSEAVC	
orf105ng-1	FRPFGLLSRAVHLNGLVES	NGRWHFWIGRR	SPHKAVDPG	KLDNIAGGGV	SGGEMPSEAVC	
	130	140	150	160	170	180
orf105-1.pep	RESSEEAGLDKTLPLIRPV	SQLHSLRSV	SRGVHNEILY	VFDAVLPETF	LPENQDGEVAG	
orf105ng-1	RESSEEAGLDKTLFPLIRPV	SRLHSLRPV	SRGVHNEILY	VFDAVLPETF	LPENQDGEVAG	
	190	200	210	220	230	240
orf105-1.pep	FEKMDIGGLLDAMLSGNM	HDAQLVTLDA	FCRYGLIDAA	HPLSEWLDGIR	LX	
orf105ng-1	FEKMDIGGLLDAMLSKNM	HDAQLVTLDA	FYRYGLIDAA	HPLSEWLDGIR	LX	
	250	260	270	280	290	

Furthermore, ORF105ng-1 shows homology with a yeast enzyme:

```

5      sp|P41888|TNR3_SCHPO THIAMIN PYROPHOSPHOKINASE (TPK) (THIAMIN KINASE)
      >gi|1076928|pir||S52350 thiamin pyrophosphokinase (EC 2.7.6.2) - fission yeast
      (Schizosaccharomyces pombe) >gi|666111 (X84417) thiamin pyrophosphokinase
      [Schizosaccharomyces pombe] >gi|2330852|gnl|PID|e334056 (Z98533) thiamin
      pyrophosphokinase [Schizosaccharomyces pombe] Length = 569
      Score = 105 bits (259), Expect = 4e-22
      Identities = 64/192 (33%), Positives = 94/192 (48%), Gaps = 3/192 (1%)

10     Query: 268 NKAGLLHGWRNECFDLTDGGGNPLFTLERAARFPFGLLSRAVHLNGLVESNGRW--HFWI 441
      N G+ WRNE + + P+ +ER F FG LS VH + + W+
      Sbjct: 96 NTFGIADQWRNELYTVYGKSKPVLAVERGGEWFLFGFLSTGVHCTMYIPATKEHPLRIWV 155

15     Query: 442 GRRSPHKAVDPGKLDNIAGGGVSGGEMPSEAVCRESEEAGLDKTLFPLIRPVSRLHSLR 621
      RRSP K P LDN GG++ G+ + +E SEEA LD + LI P + ++
      Sbjct: 156 PRRSPTKQTPWPNYLDNSVAGGIAHGDSVIGTMIKEFSEEANLDVSSMNLII-PCGTVSYIK 214

20     Query: 622 PVSRG-VHNEILYVFDVAVLPETFLPENQDGEVAGFEKMDIGGLLDAMLSKNMMHDAQLVT 798
      R + E+ YVFD + + +P DGEVAGF + + +L + K+ + LV
      Sbjct: 215 MEKRHWIQPELQYVFDLPVDDLVIPIRINDGEVAGFSLPLNQLVHELELKSFKPNCALVL 274

      Query: 799 LDAFYRYGLIDAAHP 843
      LD R+G+I HP
      Sbjct: 275 LDFLIRHGIITPQHP 289

```

25 Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 49

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 30 419>:

```

35     1 ATGAATAGAC CCAAGCAACC CTCTTCCGT CCCGAAGTCG CCGTTGCCCG
      51 CCAAACCAGC CTGACGGGTA AAGTGATTCT GACACGACCG TTGTCATTTT
      101 CCCTATGGAC GACATTTGCA TCGATATCTG CGTTATTGAT TATCCTGTTT
      151 TTGATATTTG GTAACATATAC GCGAAAGACA ACAGTGGAGG GACAAATTTT
      201 ACCTGCATCG GCGTAATCA GGGTGTATGC ACCGgATACG rGkACAATTA
      251 CAGCGAAATT CGTGAAGAT GGmsAAAAGG TTAAGGCTGG CGACAAGCTA
      301 TTTGCGCTTT CGACCTCACG TTTCGGCGCA GGAGGTAGCG TGCAGCAGCA
      351 GTTGAACACG GAGGCAGTTT TGAAGAAAAC GTTGGCAGAA CAGGAAGTGG
      401 GTCGCTCTGAA GCTGATACAC GGAATGAAA CGCGCAGCcT TAAAGCAACT
      451 GTCGAACGTT TGGAAAACCA GGAAGTCCAT ATTTGCAAC AGATAGACGG
      501 TCAGAAAAGG CGCATTAGAC TTGCGGAAGA AATGTTGCAG AAATATCGTT
      551 TCCTATCCGC .CAATGA

```

This corresponds to the amino acid sequence <SEQ ID 420; ORF107>:

```

45     1 MNRPKQFFFR PEVAVARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF
      51 LIFGNYTRKT TVEGQILPAS GVIRVYAPDT XTITAKFVED GXKVKAGDKL
      101 FALSTSFRGA GGSVQQQLKT EAVLKKTLE QELGRLKLIH GNETRSLKAT
      151 VERLENQELH ISQQIDGQKR RIRLAEMLQ KYRFLSXQ*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF107 shows 97.8% identity over a 186aa overlap with an ORF (ORF107a) from strain A of *N.meningitidis*:

-259-

	orf107.pep	MNRPKQPFRRPEVAVARQTS	LTGKVILTRPLS	FSLWTTFASISALLIILFLIFGNYTRKT	
	orf107a	MNRPKQPFRRPEVAVARQTS	LTGKVILTRPLS	FSLWTTFASISALLIILFLIFGNYTRKT	
5		10	20	30	40
	orf107.pep	70	80	90	100
	orf107a	70	80	90	100
10		110	120		
	orf107.pep	130	140	150	160
	orf107a	130	140	150	160
15		170	180		
	orf107.pep	189			
	orf107a	189			
20		190	200	210	220
	orf107.pep	190	200	210	220
	orf107a	190	200	210	220
		230			

The complete length ORF107a nucleotide sequence <SEQ ID 421> is:

25	1	ATGAATAGAC	CCAAGCAACC	NTTCTTCCGT	CCCGAAGTCG	CCGTTGCCCCG
	51	CCAAACCAGC	CTGACGGGTA	AAGTGATTCT	GACACGACCG	TTGTCATTTT
	101	CCCTATGGAC	GACATTGCA	TCGATATCTG	CGTTATTGAT	TATCCTGTTT
	151	TTGATATTG	GTAACATAC	GCGAAAGACA	ACAGTGGAGG	GACAAATTTT
	201	ACCTGCATCG	GGCGTAATCA	GGGTGTATGC	ACCGGATACG	GGGACAATTA
30	251	CNCGCAAATT	CNTGGAAGAT	GGAGAAAAGG	TTAAGGCTGG	CGACAAGCTA
	301	TTTGCGCTTT	CGACCTCAG	TTTCGGCGCA	GGAGATAGCG	TGCAGCAGCA
	351	GTTGAAAACG	GAGGCAGTTT	TGAAGAAAAC	GTTGGCAGAA	CAGGAACTGG
	401	GTCGTCTGAA	GCTGATACAC	GGGAATGAAA	CGCGCAGCCT	TAAAGCAACT
	451	GTCGAACGTT	TGGA AAAACCA	GGA ACTCCAT	ATTTCGCAAC	AGATAGACGG
35	501	TCAGAAAAGG	CGCATTAGAC	TTGCGGAAGA	AATGTTGCAG	AAATATCGTT
	551	TCCTATCCGC	CAATGATGCA	GTGCCAAAAC	AAGAAATGAT	GAATGTCAAG
	601	GCAGAGCTTT	TAGAGCAGAA	AGCCAAACTT	GATGCCTACC	GCCGAGAAGA
	651	AGTCGGGCTG	CTTCAGGAAA	TCCGCACGCA	GAATCTGACA	TTGGNNAGCC
	701	TCCCCCAAGC	GGCATGA			

This encodes a protein having amino acid sequence <SEQ ID 422>:

40	1	MNRPKQPFRR	PEVAVARQTS	LTGKVILTRP	LSFSLWTTFA	SISALLIILF
	51	LIFGNYTRKT	TVEGQILPAS	GVIRVYAPDT	GTITAKFXED	GEKVKAGDKL
	101	FALSTSRFGA	GDSVQQQLKT	EAVLKKTAE	QELGRLKLIH	GNETRSLKAT
	151	VERLENQELH	ISQQIDGQKR	RIRLAEEMLO	KYRFLSANDA	VPKQEMMNVK
45	201	AELLEQAKL	DAYRREEVGL	LQEIRTQNL	LXSLPQAA*	

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF107 shows 95.7% identity over a 188aa overlap with a predicted ORF (ORF107.ng) from *N.gonorrhoeae*:

50	orf107.pep	MNRPKQPFRRPEVAVARQTS	LTGKVILTRPLS	FSLWTTFASISALLIILFLIFGNYTRKT	60
	orf107ng	MNRPKQPFRRPEVAIARQTS	LTGKVILTRPLS	FSLWTTFASISALLIILFLIFGNYTRKT	60
	orf107.pep	TVEGQILPASGVIRVYAPDT	XTITAKFVEDG	XKVKAGDKLFALSTSRFGAGGSVQQQLKT	120
55	orf107ng	TMEGQILPASGVIRVYAPDT	GTITAKFVEDG	EKVKAGDKLFALSTSRFGAGGSVQQQLKT	120
	orf107.pep	EAVLKKTAEQELGRLKLIH	GNETRSLKAT	VERLENQELHISQQIDGQKRRIRLAEEMLO	180
60	orf107ng	EAVLKKTAEQELGRLKLIH	ENETRSLKAT	VERLENQKLHISQQIDGQKRRIRLAEEMLR	180
	orf107.pep	KYRFLSXQ	188		
	orf107ng	KYRFLSAQ	188		

The complete length ORF107ng nucleotide sequence <SEQ ID 423> is predicted to encode a protein having amino acid sequence <SEQ ID 424>:

```

      1  MNRPKQFFFR PEVAIARQTS LTGKVILTRP LSFSLWTTFA SISALLIILE
      51  LIFGNYTRKT TMEGQILPAS GVIRVYAPDT GTITAKFVED GEKVKAGDKL
5      101  FALSTSRFGA GGSVQQQLKT EAVLKKTAE QELGRLKLIH ENETRSLKAT
      151  VERLENQKLH ISQQIDGQKR RIRLAEMLR KYRFLSAQ*

```

Based on the presence of a putative ransmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 10 Example 50

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 425>:

```

      1  ATGCTGAATA CTTTTTTTGC CGTATTGGGC GGCTGCCTGC TGCT.TTGCC
      51  GTGCGGCAAA TCCGTAAATA CGGCGGTACA GCCGCAAAAC GCGGTACAAA
15     101  GCGCGCCGAA ACCGGTTTTTC AAAGTCATAT ATATCGACAA TACGGCGATT
      151  GCCGGTTTGG ATTTGGGACA AAGCAGCGAA GGCAAAACCA ACGACGGCAA
      201  AAAACAAATC AGTTATCCGA TTAAAGGCTT GCCGGAACAA AATGTTATCC
      251  GACTGATCGG CAAGCATCCC GCGGACTTGG AAGCCGTCAG CGGCAAATGT
      301  ATGGAACCG ATGATAAGGA CAGTCCGGCA GGTGGGCGAG AAAACGGCGT
20     351  GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
      401  GCAAACTGAC GGATTACCTA GTTTCGCATG CCGCCCTGCA ACCCTATCAG
      451  GCAGGCAAAA GCGGCTATGC CGCCGTGCAG AACGGACGCT ATGTGCTGGA
      501  AATCGACAGC GAAGGGGCGT TTTATTTCG CCGCCGCCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 426; ORF108>:

```

25     1  MLNTFFAVLG GCLLXLPCGK SVNTAVQPQN AVQSAPKPVF KVIYIDNTAI
      51  AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC
      101  METDDKDSPA GWAENGVCHT LFAKLVGNIA EDGGKLDYDL VSHAALQPYQ
      151  AGKSGYAAVQ NGRYVLEIDS EGAFFRRRH Y*

```

Further work revealed the following DNA sequence <SEQ ID 427>:

```

30     1  ATGCTGAAAA CATCTTTTGC CGTATTGGGC GGCTGCCTGC TGCTTGCCCG
      51  CTGCGGCAAA TCCGAAAATA CGGCGGAACA GCCGCAAAAC GCGGTACAAA
      101  GCGCGCCGAA ACCGGTTTTTC AAAGTCAAAT ATATCGACAA TACGGCGATT
      151  GCCGGTTTGG ATTTGGGACA AAGCAGCGAA GGCAAAACCA ACGACGGCAA
35     201  AAAACAAATC AGTTATCCGA TTAAAGGCTT GCCGGAACAA AATGTTATCC
      251  GACTGATCGG CAAGCATCCC GCGGACTTGG AAGCCGTCAG CGGCAAATGT
      301  ATGGAACCG ATGATAAGGA CAGTCCGGCA GGTGGGCGAG AAAACGGCGT
      351  GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
      401  GCAAACTGAC GGATTACCTA GTTTCGCATG CCGCCCTGCA ACCCTATCAG
40     451  GCAGGCAAAA GCGGCTATGC CGCCGTGCAG AACGGACGCT ATGTGCTGGA
      501  AATCGACAGC GAAGGGGCGT TTTATTTCG CCGCCGCCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 428; ORF108-1>:

```

45     1  MLKTSFAVLG GCLLLAACGK SENTAEQPQN AVQSAPKPVF KVKYIDNTAI
      51  AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC
      101  METDDKDSPA GWAENGVCHT LFAKLVGNIA EDGGKLDYDL VSHAALQPYQ
      151  AGKSGYAAVQ NGRYVLEIDS EGAFFRRRH Y*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF108 shows 88.4% identity over a 181aa overlap with a predicted ORF (ORF108.ng) from *N.gonorrhoeae*:

```

5      orf108.pep  MLNTFFAVLGGCLLXLPCGKSVNTAVQPQNAVQSAPKPVFKVIYIDNTAIAGLDLGQSSE  60
      orf108ng    MLKIPFAVLGGCLLLAACGKSENTAEQPQNAAQSAPKPVFKVKYIDNTAIAGLALGQSSE  60

10     orf108.pep  GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSPAGWAENGVCHT  120
      orf108ng    GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT  120

15     orf108.pep  LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH  181
      orf108ng    LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH  181

```

ORF108-1 shows 92.3% identity with ORF108ng over the same 181 aa overlap:

```

20     orf108-1.pep MLKTSFAVLGGCLLLAACGKSENTAEQPQNAVQSAPKPVFKVKYIDNTAIAGLDLGQSSE  60
      orf108ng-1    MLKIPFAVLGGCLLLAACGKSENTAEQPQNAAQSAPKPVFKVKYIDNTAIAGLALGQSSE  60

25     orf108-1.pep GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSPAGWAENGVCHT  120
      orf108ng-1    GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT  120

30     orf108-1.pep LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH  181
      orf108ng-1    LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH  181

```

The complete length ORF108ng nucleotide sequence <SEQ ID 429> is:

```

30      1  ATGCTGAAAa  tacctTTTGC  CGTGTtgggc  ggCtgcctGC  TGCTTGCCGC
      51  CTGCGGCAAA  TCCGAAAATa  cggcggaACA  GCCGCAAAT  gcggCACAAa
      101 GCGCGCCGAA  ACCGTTTTTC  AAAGTCAAAT  ACATCGACAA  TACGGCGATT
      151 GCCGGTTTGG  CTTTGGGACA  AAGTAGCGAA  GGCAAACCA  acgacgGCAA
      201 AAAACAAATC  AGTTATccgA  TTAAAGGCTT  GCCGGAACAA  Aacgcggtcc
      251 gGCTGACCGG  AAAGCATCCC  AACGACTTGG  AagccgtcgT  CGGCAAATGT
      35  301  ATGGAACCG  ACGGAAAGGA  CGCGCCTTCG  GGCTGGGCGG  AAAACGGCGT
      351 GTGCCATACC  TTGTTTGCCA  AACTGGTGGG  CAATATCGCC  GAAGACGGCG
      401 GCAAACGAC  TGATTACCTG  ATTCGCATT  CCGCCCTGCA  ACCCTATCAG
      451 GCAGGCAAAA  GCGGCTATGC  CGCCGTGCAG  AACGGACGCT  ATGTGCTGGA
      501 AATCGACAGC  GagggGGCGT  TTTATttccg  ccgccgcat  tattgA

```

40 This encodes a protein having amino acid sequence <SEQ ID 430>:

```

      1  MLKIPFAVLG  GCLLLAACGK  SENTAEQPQN  AAQSAPKPVF  KVKYIDNTAI
      51  AGLALGOSSE  GKTNDGKKQI  SYPIKGLPEQ  NAVRLTGKHP  NDLEAVVGKC
      101  METDGKDAPS  GWAENGVCHT  LFAKLVGNI  AEDGGKLTDL  ISHSALQPYQ
      151  AGKSGYAAVQ  NGRYVLEIDS  EGAFYFRRRH  Y*

```

45 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) and a putative ATP/GTP-binding site motif A (P-loop, double-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 50 Example 51

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 431>:

-262-

1 ATGGAAGATT TATATATAAT ACTCGCTTTG GGTTCGGTTG CGATGATTGC  
 51 CGGATTTATC GATgcatTg cGggCGGGGG TGGTTTGATT ACGCTGCCCCG  
 101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG  
 151 CTGCAAGCAG CCGCTGCTAC GTTTTCAGCT ACGGTTTCTT TTGCACGCAA  
 5 201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG  
 251 TAGGCGGCGT GGcCGGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT  
 301 CTgCTgGCGG TCGTGCCGGT TTTGTTGATA TTTGTGCGAC TGTATTTTGT  
 351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT  
 401 TTTTCTGTT cGGGCTGACG GTCGC.ACCG CTTTGGGTT TTTACGACGG  
 10 451 TGTGTTCCGA CCGGTGTGCG GCTCGTTTTT TCTGATTGCC TTTATTGTTT  
 501 TGCTCGGCTG CAAGCTGTTG AACGCGATGT CTTACACCAA ATTGGCGAAC  
 551 GTTGCTGCA ATCTTGTTT GCTATCGGTA TTCCTGCTGC ACGGTTTCGAT  
 601 TATTTTCCCG ATTGCGGCAA CGaTGCGGT CCGTGCGTTT GTCGGtCGCA  
 651 ATTTAgGTGC GAGATTTGCC GTaCgctTCG GTTCGAAGCT GATTAA

15 This corresponds to the amino acid sequence <SEQ ID 432; ORF109>:

1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK  
 51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVSLVSKDI  
 101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFLGT VXTAFGFLRR  
 151 CVRTGCRLVF SDCLYCFARL QAVERDVLHQ IGERCLQSWF AIGIPAARFD  
 20 201 YFPDCGNDGG RCVRCFEFR EICRTLRFEA D\*

Further work revealed the following DNA sequence <SEQ ID 433>:

1 ATGGAAGATT TATATATAAT ACTCGCTTTG GGTTCGGTTG CGATGATTGC  
 51 CGGATTTATC GATGCGATTG CGGGCGGGGG TGGTTTGATT ACGCTGCCCCG  
 101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG  
 25 151 CTGCAAGCAG CCGCTGCTAC GTTTTCAGCT ACGGTTTCTT TTGCACGCAA  
 201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG  
 251 TAGGCGGCGT GGCCGGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT  
 301 CTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTGCGAC TGTATTTTGT  
 351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT  
 30 401 TTTTCTGTT CGGGCTGACG GTCGCACCGC TTTTGGGTTT TTACGACGGT  
 451 GTGTCGGAC CGGGTGTGCG CTCGTTTTTT CTGATTGCCT TTATTGTTTT  
 501 GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAAA TTGGCGAACG  
 551 TTGCTGCAA TCTTGTTTCG CTATCGGTAT TCCTGCTGCA CGGTTTCGATT  
 601 ATTTTCCCGA TTGCGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA  
 35 651 TTTAGGTGCG AGATTTGCCG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC  
 701 TGCTGATTGT CATCAGCATT TCGATGGCTG TGAAATTGTT GATAGACGAG  
 751 AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTAA

This corresponds to the amino acid sequence <SEQ ID 434; ORF109-1>:

1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK  
 40 51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVSLVSKDI  
 101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFLGT VAPLLGFYDG  
 151 VFPGVGSFF LIAFIVLLGC KLLNAMSYTK LANVACNLGS LSVFLLHGS  
 201 IFPIAATMAV GAFVGANLGA REAVRFGSKL IKPLLVISI SMAVKLLIDE  
 251 RNPLYQMIVS MF\*

45 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF109 shows 95.9% identity over a 147aa overlap with an ORF (ORF109a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
50	orf109.pep	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA					
	orf109a	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA					
		10	20	30	40	50	60
55	orf109.pep	TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
	orf109a	TVSFARKGLIDWKKGLPIAAASFAGGVGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
		70	80	90	100	110	120

-263-

```

              130      140      150      160      170      180
orf109.pep    KLDGSKEGKARMSFFLFGLTVXTAFGFLRRRCVRTGCRLVFSDCLYCFARLQAVRDVLHQ
              |||||
5  orf109a     KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
              130      140      150      160      170      180

```

The complete length ORF109a nucleotide sequence <SEQ ID 435> is:

```

1  ATGGAAGATT TATACATAAT ACTCGCTTTG GGTTCGGTTG CGATGATTGC
10 51 CGGATTTATC GATGCGATTG CGGGTGGGGG TGGTTTGATT ACGCTGCCTG
101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
151 CTGCAAGCAG CCGCTGCTAC GTTTTCGGCT ACGGTTTCTT TTGCACGCAA
201 AGGTTTGATT GATTGGAAGA AAGGTCCTCC GATTGCGGCA GCATCGTTTG
251 CAGGCGGCGT GGTCCGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
301 CTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTCGCGC TGTATTTTGT
15 351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
401 TTTTCTGTT CCGTCTGACG GTTGCCACCAC TTTTGGGTTT TTACGACGGT
451 GTGTTCCGAC CGGGTGTCGG CTCGTTTTTT CTGATTGCCT TTATGTTTTT
501 GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAA TTGGCGAACG
551 TTGCTTGCAA TCTTGGTTCG CTATCGGTAT TCCTGCTGCA CGGTTCCGATT
20 601 ATTTTCCCGA TTGCGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA
651 TTTAGGTGCG AGATTGCGG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
701 TGCTGATTGT CATCAGCATT TCGATGGCTG TGAAATTGTT GATAGACGAG
751 AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 436>:

```

25 1  MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
51  LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYD
151 VFGPGVGSFF LIAFIVLLGC KLLNAMS YTK LANVACNLGS LSVFLLHGS I
30 201 IFPIAATMAV GAFVGANLGA RFAVRFGSKL IKPLLIVISI SMAVKLLIDE
251 RNPLYQMIVS MF*

```

ORF109a and ORF109-1 show 99.2% identity in 262 aa overlap:

```

              10      20      30      40      50      60
orf109a.pep    MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
35 orf109-1     MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
              10      20      30      40      50      60

              70      80      90      100     110     120
orf109a.pep    TVSFARKGLIDWKKGLPIAAASFAGGVGALS VSVLVS KDILLAVVPVLLIFVALYFVFSP
40 orf109-1     TVSFARKGLIDWKKGLPIAAASFVGGVAGALS VSVLVS KDILLAVVPVLLIFVALYFVFSP
              70      80      90      100     110     120

              130     140     150     160     170     180
orf109a.pep    KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
45 orf109-1     KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
              130     140     150     160     170     180

              190     200     210     220     230     240
orf109a.pep    LANVACNLGSLSVFLLHGSIIIFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
50 orf109-1     LANVACNLGSLSVFLLHGSIIIFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
              190     200     210     220     230     240

              250     260
orf109a.pep    SMAVKLLIDERNPLYQMIVSMFX
55 orf109-1     SMAVKLLIDERNPLYQMIVSMFX
              250     260
60

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF109 shows 98.3% identity over a 231aa overlap with a predicted ORF (ORF109.ng) from *N.gonorrhoeae*:

```

5      orf109.pep  MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA  60
      orf109ng    MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA  60

10     orf109.pep  TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP  120
      orf109ng    TVSFARKGLIDWKKGLPIAAASFAGGVGALSVSLVSKDILLAVVPVLLIFVALYFVFSP  120

15     orf109.pep  KLDGSKEGKARMSFFLFGLTVXTAFGFLRRCVRTGCRLVFSDCLYCFARLQAUERDVLHQ  180
      orf109ng    KLDGSKEGKARMSFFLFGLTVATAFGFLRRCVRTGCRLVFSDCLYCFARLQAUERDVLHQ  180

      orf109.pep  IGERCLQSWFAIGIPAARFDYFPDCGNDGGRCVCRCFRCEICRTLRFED  231
      orf109ng    IGERCLQSWFAIGIPAARFDYFPDCGNDGGRCVCRCFRCEICRPLRFED  231

```

20 An ORF109ng nucleotide sequence <SEQ ID 437> was predicted to encode a protein having amino acid sequence <SEQ ID 438>:

```

25      1  MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
      51  LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVGA LSVSLVSKDI
      101  LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VATAFGFLRR
      151  CVRTGCRLVF SDCLYCFARL QAUERDVLHQ IGERCLQSWF AIGIPAARFD
      201  YFPDCGNDGG RVCRCFRCE EICRPLRFEA D*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 439>:

```

30      1  ATGGAAGATT TATACATAAT ACTCGCTTTG GGTTTGGTTG CGATGATCGC
      51  CGGATTTATC GATGCGATTG CGGGCGGGGG TGGTTTGATT ACGCTGCCTG
      101  CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
      151  CTGCAAGCAG CCGCTGCTAC GTTTTCGGCT ACGGTTTCTT TTGCACGCAA
      201  AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
      251  CAGGCGGCGT GGTGCGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
      301  TTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTCGCGC TGTATTTTGT
      351  GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
      401  TTTTCTATT  CGGGCTGACG GTTGACCGC TTTTGGGTTT TTACGACGGT
      451  GTGTTGCGAC CGGGTGTCGG CTCGTTTTTT CTGATTGCCT TTATTGTTTT
      501  GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAA TTGGCGAACG
      551  TTGCTTGCAA TCTTGGTTCG CTATCGGTAT TCCTGCTGCA CGGTTTCGATT
      601  ATTTTCCCGA TTGTGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA
      651  TTTAGGTGCG AGATTGCGG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
      701  TGCTGATTGT CATCAGCATT TCGATGGCTG TGAAATTGTT GATAGACGAG
      751  AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTTAA

```

This corresponds to the amino acid sequence <SEQ ID 440; ORF109ng-1>:

```

45      1  MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
      51  LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVGA LSVSLVSKDI
      101  LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG
      151  VFGPGVGSFF LIAFIVLLGC KLLNAMS YTK LANVACNLGS LSVFLLHGS
      201  IFPIVATMAV GAFVGANLGA REAVRFGSKL IKPLLIVISI SMAVKLLIDE
      251  RNPLYQMIVS MF*

```

50 ORF109ng-1 and ORF109-1 show 98.9% identity in 262 aa overlap:

```

55      10      20      30      40      50      60
      orf109ng-1.pep MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
      orf109-1      MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA

      70      80      90      100     110     120
      orf109ng-1.pep TVSFARKGLIDWKKGLPIAAASFAGGVGALSVSLVSKDILLAVVPVLLIFVALYFVFSP

```



-265-

```

      |||:|||||
orf109-1 TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
      70      80      90      100      110      120

5      130      140      150      160      170      180
orf109ng-1.pep KLDGSKEGKARMSFFLFGTLVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
      |||:|||||
orf109-1 KLDGSKEGKARMSFFLFGTLVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
      130      140      150      160      170      180

10     190      200      210      220      230      240
orf109ng-1.pep LANVACNLGSLSVFLLHGSIIFFPIVATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
      |||:|||||
orf109-1 LANVACNLGSLSVFLLHGSIIFFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
      190      200      210      220      230      240

15     250      260
orf109ng-1.pep SMAVKLLIDERNPLYQMIVSMFX
      |||:|||||
20     orf109-1 SMAVKLLIDERNPLYQMIVSMFX
      250      260

```

In addition, ORF109ng-1 shows homology to a hypothetical *Pseudomonas* protein:

```

25     sp|P29942|YCB9_PSEDE HYPOTHETICAL 27.4 KD PROTEIN IN COBO 3'REGION (ORF9)
>gi|94984|pir|I38164 hypothetical protein 9 - Pseudomonas sp >gi|551929
(M62866) ORF9 [Pseudomonas denitrificans] Length = 261
Score = 175 bits (439), Expect = 3e-43
Identities = 83/214 (38%), Positives = 131/214 (60%), Gaps = 1/214 (0%)

30     Query: 41 PPVSAIATNKLQXXXXXXXXXXXXXKRLIDWKKGLPIXXXXXXXXXXXXXXXXXKDI 100
PP+ + TNKLQ R+G ++ K+ LP+ D+
Sbjct: 43 PPLQTLGTNKLQGLFGSGSATLSYARRGHVNLKEQLPMALMSAAGAVLGALLATIVPGDV 102

Query: 101 LLAVVPVLLIFVALYFVFSPKLDGSKEGKARMSFFLFGTLVAPLLGFYDGVFGPGVGSFF 160
L A++P LLI +ALYF P + G + +R++ F+F LT+ PL+GFYDGVFGPG GSFF
35     Sbjct: 103 LKAILPFLIIAIALYFGLKPNM-GDVDQHSRVTPEVFTLTLVPLIGFYDGVFGPGTGSFF 161

Query: 161 LIAFIVLLGCKLLNAMS YTKLANVACNLGSLSVFLLHGSIIFFPIVATMAVGAFVGANLGA 220
++ F+ L G +L A ++TK N N+G+ VFL G++++ + M +G F+GA +G+
40     Sbjct: 162 MLGFVTLAGFGVLKATAHTKFLNFGSNVGAFGVFLFFGAVLWKVGLLMGLGQFLGAQVGS 221

Query: 221 RFAVRFGSKLIKPLLIVISISMAVKLLIDERNPL 254
R+A+ G+K+IKPLL+++SI++A++LL D +PL
Sbjct: 222 RYAMAKGAKIIKPLLIVISIALAIRLLADPTHPL 255

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 52

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 441>:

```

50      1 ..CTGCTAGGGT ATTGCATCGG TTATCGGTAC GGCTGTTGCA GCAAACCAG
51      CCGCAGACGG ATTATTGGT CAAATTCGGA TCGTTTGGG CGAG.ATTTT
101     TGGTTTCTG GGACTGTATG ACGTCTATGC TTCGGCATGG TTTGTCGTTA
151     TCATGATGTT TTTGGTGGT TCTACAGTT TGTGCCTGAT TCGCAATGTG
201     CCGCCGTCT GCGCGAAAT GAAGTCTTTT CGGGAAAAGG TTAAGAAAA
55     251 ATCTCTGGCG GCGATGCGCC ATTCTTCGCT GTTGATGTA AAAATTGCGC
301     CCGAGGTTGC CAAACGTTAT CTGGAAGTAC AAGGTTTTC GGGGAAACC
351     ATTAACCGTG AAGACGGGTC GGTCTGATT GCCGCCAAA AAGGCACAAT
401     GAACAAATGG GGCTATATCT TTGCCATGT TGCTTTGATT GTCATTGACC
451     TGGGCGGTT GATAGACAGT AACCTGCTGT TGAAACTGGG TATGCTGACC
60     501 GGTGCGATTG TTCCGGACAA TCAGGCGGTT TATGCCAAGG ATTC.AAGC

```

-266-

551 CCGAAAGTAT .TTTGGGTGC gTCCAATCTC TCATTAGGG GCAACGTCAA  
601 TATTTCCG.A GGGGCAGAgT GCGGATGTGG TTTTCCTGA

This corresponds to the amino acid sequence <SEQ ID 442; ORF110>:

5           1   ..LLGIASVIGT LLQQNQPTD YLVKFGSFWA XIFGFLGLYD VYASAWFVVI  
          51   MMFLVVSTSL CLIRNVPPFW REMKSFREKV KEKSLAAMRH SSLLDVKIAP  
         101   EVAKRYLEVQ GFQGTINRE DGSVLIAAK GTMNKGYIF AHVALIVICL  
         151   GGLIDSNLLL KLGMLTGRIF RTIRREMPRI XKPESXFGCV QSLI\*GQRQY  
         201   FXRGRVRMWF S\*

Computer analysis of this amino acid sequence gave the following results:

#### 10 Homology with ORF88a from *N.meningitidis* (strain A)

ORF110 shows 91.5% identity over a 188aa overlap with ORF88a from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf88a.pep	MSKSRRSP	LLSRPWFAFFSSMRFA	ALLSLLGIASVIGT	VLQQNQPTD	YLVKFGSFWA	
15	orf110			LLGIASVIGT	LLQQNQPTD	YLVKFGSFWA	
				10	20	30	
		70	80	90	100	110	120
20	orf88a.pep	QIFGFLGLYDVYASAWFV	VIMMFLVVSTSLCLIRNVPPFW	REMKSFREKVKEKSLAAMRH			
	orf110	XIFGFLGLYDVYASAWFV	VIMMFLVVSTSLCLIRNVPPFW	REMKSFREKVKEKSLAAMRH			
		40	50	60	70	80	90
		130	140	150	160	170	180
25	orf88a.pep	SSLLDVKIAPEVAKRYLEVQ	GFQGTINREDGSVLIAAK	GMTMNKGYIFAHVALIVICL			
	orf110	SSLLDVKIAPEVAKRYLEVQ	GFQGTINREDGSVLIAAK	GMTMNKGYIFAHVALIVICL			
		100	110	120	130	140	150
30	orf88a.pep	GGLIDSNLLLKLGMLTGRIV	PDNQAVYAKDFKPESILGAS	NLSFRGNVNISEGQSADVVF			
	orf110	GGLIDSNLLLKLGMLTGRIF	RTIRREMPRIKKPESXFGCV	QSLIXGQRQYFXRGRVRMWF			
35		160	170	180	190	200	210
	orf88a.pep	LNADNGILVQDLPFEVKL	KKFHIDFYNTGMPRDFAS	DIETDKATGEKLERTIRVNHPLT			
	orf110	SX					

40 However, ORF88 and ORF110 do not align, because they represent two different fragments of the same protein.

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF110 shows 88.6% identity over a 211aa overlap with a predicted ORF (ORF110.ng) from *N. gonorrhoeae*:

45	orf110.pep		LLGIASVIGTLLQQNQPTD	YLVKFGSFWA	30
	orf110ng	MSKSRI	SPTLLSRPWFAFFSSMRFA	ALLSLLGIASVIGT	VLQQNQPTD
					60
50	orf110.pep	XIFGFLGLYDVYASAWFV	VIMMFLVVSTSLCLIRNVPPFW	REMKSFREKVKEKSLAAMRH	90
	orf110ng	RIFDFLGLYDVYASAWFV	VIMMFLVVSTSLCLIRNVPPFW	REMKSFREKVKEKSLAAMRH	120
	orf110.pep	SSLLDVKIAPEVAKRYLEVQ	GFQGTINREDGSVLIAAK	GMTMNKGYIFAHVALIVICL	150
55	orf110ng	SSLLDVKIAPEVAKRYLEV	RGFQGTVSREDGSVLIAAK	GMTMNKGYIXAHVALIVICL	180

```

orf110.pep  GGLIDSNLLKLGMLTGRIFRTIRRFMPRIKKPESXFGCVQSLIXGQRQYFXRGRVRMWF  210
| ||: |||||: ||||: || |||| |||| : ||||| ||||| ||: |||||
orf110ng    GRLINXNLLKLGMLAGSIFRNNRRVMPRIKPSIWWGGVQSLIKGQRQYFQRGKVRMWF  240
5          orf110.pep  S  211
              |
          orf110ng    S  241

```

The complete length ORF110ng nucleotide sequence <SEQ ID 443> is predicted to encode a protein having amino acid sequence <SEQ ID 444>:

```

10          1  MSKSRISPTL LSRPWFAFFS SMRFAVALLS LLGIASVIGT VLQQNQPTD
          51  YLVKFGPFWT RIFDFLGLYD VYASAWFVVI MMFLVVSTSL CLIRNVPPFW
          101  REMKSFREKV KEKSLAAMRH SLLDVKIAP EVAKRYLEVR GFQKTVSRE
          151  DGSVLIAAKK GTMNKWGYIX AHVALIVICL GRLINXNLLL KLGMLAGSIF
          201  RNNRRVMPRI SKPESIWWGGV QSLIKGQRQY FQRGKVRMWF S*

```

15 Based on the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 53

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 445>:

```

20          1  ATGCCGTCTG AAACACGCCT GCCGAAC TTT ATCCGCGTCT TGATATTTGC
          51  CCTGGGTTC ATCTTCTCTGA ACGCCTGTTT GGAACAAACC GCGCAAACCG
          101  TTACCCTGCA AGGCGAAACG ATGGGCACGA CCTATACCGT CAAATACCTT
          151  TCAAATAATC GGGACAAACT CCCCTCACCT GCCGAAATAC AAAAACGCAT
          201  CGATGACGCG CTTAAAGAAG TCAACCGGCA GATGTCCACC TATCAGCCCG
25          251  ACTCCGAAAT CAGCCGGTTC AACCAACACA CAGCCGGCAA GCCCTCCGC
          301  ATTTCAAGCG ACTTCGCACA CGTTACTGCC GAAGCCGTCC GCCTGAACCG
          351  CCTGACACAC GCGCGCTGG ACGTAACCGT CCGCCCCCTTG GTCAACCTTT
          401  GGGGATTTCG CCCCACAAA TCCGTTACCC GTGAACCGTC GCCGGAACAA
          451  ATCAAAACAG CCGCATCTTA TACGGGCATA GACAAAATCA TTTGAAACA
30          501  AGGCAAAGAT TACGCTTCTT TGAGCAAAAC CCACCCCAAG GCCTATTTGG
          551  ATTTATCTTC GATTGCCAAA GGCTTCGGCG TTGATAAAGT TCGGGGCGAA
          601  CTGGAATAAT ACGGCATTCA AAATATCTG GTCGAAATCG GCGGCGAGTT
          651  GCACGGCAAA GGCAAAAACG CGCGCGGCGA ACCGTGGCGC ATCGGTATCG
          701  AGCAGCCCAA TATCGTCCAA GGCGGCAATA CGCAGATTAT CGTCCCGCTG
35          751  AACAACCGTT CGCTTGCCAC TTCCGGCGAT TACCGTATTT TCCACGTCGA
          801  TAAAAACGGC AAACGCCTCT CCCATATCAT CAACCCGAAC AACAAACGAC
          851  CCATCAGCCA CAACCTCGCC TCCATCAGCG TGGTCGCAGA CAGTGCATG
          901  ACGGCGGACG GCTTGTCAC AGGATTATTC GTATTGGGCG AAACCGAAGC
          951  CTTAAAGCTG GCAGAGCGCG AAAAATCGC TGTTTTCCTG ATTGTCAGGG
40          1001  ATAAAGGCGG CTACCGCACC GCCATGTCTT CCGAATTGTA AAAACTGCTC
          1051  CGCTAA

```

This corresponds to the amino acid sequence <SEQ ID 446; ORF111>:

```

45          1  MPSETRLPNF IRVLIFALGF IFLNACSEQT AQTVTLQGET MGTTYTVKYL
          51  SNNRDKLPSP AEIQKRIDDA LKEVNRQMSY YQPDSEISRF NQHTAGKPLR
          101  ISSDFAHVTA EAVRLNRLTH GALDVTGVL VNLWGFDPK SVTREPSPEQ
          151  IKQAASYTGI DKIIKQKGD YASLSKTHPK AYLDLSSIAK GFGVDKVAGE
          201  LEKYGIQNYL VEIGGELHGK GKNARGEPR IGIEQPNIVQ GGNTQIIIVPL
          251  NNRSLATSGD YRIFHVDKNG KRLSHIINPN NKRPISHNLA SISVVDASAM
          301  TADGLSTGLF VLGETEALKL AEREKLAVFL IVRDKGGYRT AMSSEFEKLL
50          351  R*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF111 shows 96.9% identity over a 351aa overlap with an ORF (ORF111a) from strain A of *N. meningitidis*:

5	orf111a.pep	MPSETRLPNFIRTLIFALSIFLNACSEQTAQTVTLQGETMGTTYTVKYLSNNRDXLPSP
	orf111	MPSETRLPNFIRVLIFALGFIFLNACSEQTAQTVTLQGETMGTTYTVKYLSNNRDKLPSP
10	orf111a.pep	AEIQXRIDDALKEVNRQMSTYQPDSEISRNFQHTAGKPLRISSDFAHVTAEAVHLNRLTH
	orf111	AEIQXRIDDALKEVNRQMSTYQPDSEISRNFQHTAGKPLRISSDFAHVTAEAVRLNRLTH
15	orf111a.pep	GALDVTVGPLVNLWGFGPDKSVTREPSPEQIKQAASYTGIDKIILKQKDYASLSKTHPK
	orf111	GALDVTVGPLVNLWGFGPDKSVTREPSPEQIKQAASYTGIDKIILKQKDYASLSKTHPK
25	orf111a.pep	AYLDLSSIAGFGVDXVAGELEKYGIQNYLVEIGGELHGKXKNARGEPPWRIGIEQPNIVQ
	orf111	AYLDLSSIAGFGVDKXVAGELEKYGIQNYLVEIGGELHGKGNARGEPPWRIGIEQPNIVQ
30	orf111a.pep	GGNTQIIIVPLNNRSXATSGDYRIFHVDKSGKRLSHIINPNNKRPISHNLAISISVXADSAM
	orf111	GGNTQIIIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNNKRPISHNLAISISVVADSAM
35	orf111a.pep	TADGXSTGLFVLGETEALKLAEREKLAVFLIVRDKGGYRTAMSSEFEKLLRX
	orf111	TADGLSTGLFVLGETEALKLAEREKLAVFLIVRDKGGYRTAMSSEFEKLLRX

The complete length ORF111a nucleotide sequence <SEQ ID 447> is:

40	1	ATGCCGTCG	AAACACGCCT	GCCGAACCTT	ATCCGCACCT	TGATATTTGC
	51	CCTGAGTTT	ATCTTCCTGA	ACGCCTGTTC	GGAACAAACC	GCGCAAACCG
	101	TTACCCTGCA	AGGTGAAACG	ATGGGCACGA	CCTATACCGT	CAAATACCTT
	151	TCAAATAATC	GGGACNAACT	CCCNACACCT	GCCGAAATAC	AAAANCGCAT
	201	CGATGACGCG	CTTAAAGAAG	TCAACCGGCA	GATGTCCACC	TATCAGCCCG
45	251	ACTCCGAAAT	CATCCGGTTC	AACCAACACA	CAGCCGGCAA	GCCCCCTCCG
	301	ATTTCAAGCG	ACTTCGCACA	CGTTACTGCC	GAAGCCGTCC	ACCTGAACCG
	351	CCTGACACAC	GGCGCGCTGG	ACGTAACCGT	CGGCCCTTG	GTCAACCTTT
	401	GGGGATTCCG	CCCCGACAAA	TCCGTTACCC	GTGAACCGTC	GCCGGAACAA
	451	ATCAAACAAG	CAGCATCTTA	TACGGGCATA	GACAAAATCA	TTTTGAAACA
50	501	AGGCAAGAT	TACGTTCTCT	TGAGCAAAAC	CCACCCCAAG	GCCTATTTGG
	551	ATTTATCTTC	GATTGCCAAA	GGCTTCGGCG	TTGATNANGT	TGCGGGCGAA
	601	CTGGAAAAAT	ACGGCATTCA	AAATTATCTG	GTGAAATCG	GCGGNGAGTT
	651	GCACGGCAAA	GNCAAAAACG	CGCGCGGCGA	ACCTTGGCGC	ATCGGCATCG
	701	AACAGCCCAA	CATCGTCCAA	GGCGGCAATA	CGCAGATTAT	CGTCCCCTG
55	751	AACAACCGTT	CGNTTGCCAC	TTCCGGCGAT	TACCGTATTT	TCCACGTCGA
	801	TAAAAGCGGC	AAACGCCTCT	CCCATATCAT	TAATCCGAAC	AACAAACGAC
	851	CCATCAGCCA	CAACCTCGCC	TCCATCAGCG	TGNTCGCAGA	CAGTGCGATG
	901	ACGGCGGACG	GCTTNTCCAC	AGGATTATTC	GTATTGGGCG	AAACCGAAGC
	951	CTTAAAGCTG	GCAGAGCGCG	AAAACTCGC	TGTTTTCCTG	ATTGTCAGGG
60	1001	ATAAAGCGG	CTACCGCACC	GCCATGTCTT	CCGAATTTGA	AAAACTGCTC
	1051	CGCTAA				

This encodes a protein having amino acid sequence <SEQ ID 448>:

1 MPSETRLPNF IRTLIFALS IFLNACSEQT AQTVTLQGET MGTTYTVKYL  
51 SNNRDXLPSP AEIQXRIDDA LKEVNRQMST YQPDSEISRF NQHTAGKPLR

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF111 shows 96.6% identity over a 351aa overlap with a predicted ORF (ORF111.ng) from *N.*

10 *gonorrhoeae:*

		10	20	30	40	50	60
	orf111ng	MPSETRLPNLIRALIFALGFIFLNACSEQTAQTVTLQGETMGTTYTVKYLSNNRDKLPSP      ::   :					
15	orf111	MPSETRLPNFIRVLIFALGFIFLNACSEQTAQTVTLQGETMGTTYTVKYLSNNRDKLPSP 10 20 30 40 50 60					
		70	80	90	100	110	120
	orf111	AKIQKRIDDALKEVNRQMSTYQTDSEISRFNQHTAGKPLRISSDFAHVTAEA VRLNRLTH  :     :     :					
20	orf111	AETQRIDDALEKNRMSTYQPDSEISRFNQHTAGKPLRISSDFAHVTAEA VRLNRLTH 70 80 90 100 110 120					
		130	140	150	160	170	180
25	orf111ng	GALDVTVGPLVN LWGFGPKSVTREPSPEQIKQAASYTGIDKIILQQGKD YASLSKTHPK      :     :					
	orf111	GALDVTVGPLVN LWGFGPKSVTREPSPEQIKQAASYTGIDKIILKQGKD YASLSKTHPK 130 140 150 160 170 180					
		190	200	210	220	230	240
30	orf111ng	AYLDLSSI AKGFGVDKVAGELEKYGI QNYLVEIGGELHGKGKN AGEPWRIGIEQPNI IQ      :     :					
	orf111	AYLDLSSI AKGFGVDKVAGELEKYGI QNYLVEIGGELHGKGKN ARGE PWRIGIEQPNI VQ 190 200 210 220 230 240					
		250	260	270	280	290	300
35	orf111ng	GGNTQIIIVPLNN RSLATSGDYRIFHVD KNGKRLSHIINPNNK RPISHNLASI SVVS DSAM      :     :					
	orf111	GGNTQIIIVPLNN RSLATSGDYRIFHVD KNGKRLSHIINPNNK RPISHNL ASI SVVA DSAM 250 260 270 280 290 300					
40		310	320	330	340	350	
	orf111ng	TADGLSTGL FVLGETEALRLAEQEKLAV FLIVRD KDGYRTAMS SEFAKLLRX      :     :					
	orf111	TADGLSTGL FVLGETEALKLAEREKLAV FLIVRD KG GYRTAMS SEFEKLLRX 310 320 330 340 350					
45							

The complete length ORF111ng nucleotide sequence <SEQ ID 449> is:

	1	ATGCCGCTCTG	AAACACGCCT	GCCGAACCTT	ATCCGCGCCT	TGATATTTGC
	51	CCTGGGTTC	ATCTTCTCTGA	ACGCCTGTTC	GGaacaacC	GCGCAaaccg
50	101	TTACCCTGCA	AGGCGAAA <sub>cg</sub>	aTGGGTACGA	CCTATACCGT	CAAATACCTT
	151	TCAAATAATC	GGGACAAAT	CCCCTCCCCT	GCCAAAAATAC	AAAAGCGCAT
	201	TGATGATGCG	CTTAAAGAAG	TCAACCGGCA	GATGTCCACC	TACCAGACCG
	251	ATTTCCGAAAT	GACCGGTTT	AACCAACACA	CAGCCGGCAA	GCCCCCTCCG
	301	ATTTCAAGCG	ATTTTCGCACA	CGTTACCGCC	GAAGCCGTCC	GCCTGAACCG
55	351	CCTGACTCAC	GGCGCACTGG	ACGTAACCGT	CGGCCCTTTG	GTCAACCTTT
	401	GGGGGTTTCGG	CCCCGACAAA	TCCGTTACCC	GTGAACCGTC	GCCGGAACAA
	451	ATCAAAACAGG	CGGCATCTTA	TACGGGCATA	GACAAAATCA	TTTTCGAACA
	501	AGGCAAAGAT	TACGCTTCCT	TGAGCAAAAC	CCACCCCAA	GCCTATTTGG
	551	ATTTATCTTC	GATTGCCAAA	GGCTTCGGCG	TTGATAAAGT	TGCGGGCGAA
60	601	CTGGAAAAAT	ACGGCATTTCA	AAATTATCTG	GTGGAAT <sub>cg</sub>	g <sub>cg</sub> gcGAGTT
	651	GCACGGC <sub>AAA</sub>	GGCAAAAAAT	CGCACGGCGA	ACCGTGGCGC	ATCGGTATAG
	701	AGCAACCCAA	TATCATCCAA	G <sub>cg</sub> gcGA <sub>ata</sub>	CGCAGATTat	cgtcccgc <sub>tg</sub>
	751	aaCaaccg <sub>tt</sub>	cgctTGCCAC	TTCCGGCGAT	TAccg <sub>ta</sub> TTT	tccacg <sub>tc</sub> gA
	801	TAAAAAcg <sub>gc</sub>	aaacgcct <sub>tt</sub>	cccacaTCAT	CAATCCC <sub>a</sub> AC	aacAAACg <sub>ac</sub>
65	851	ccATCAGC <sub>ca</sub>	caacctcg <sub>cc</sub>	tccatcag <sub>cg</sub>	tggtctc <sub>AG</sub> A	CAGTGC <sub>AA</sub> TG
	901	ACGGCGGACG	GTTt <sub>at</sub> CCAC	AGGATTATTT	GTTTTAGGCG	AAACCGAATG
	951	CTTAAGGCTG	GCAGAA <sub>CA</sub> AG	AAAAACTCGC	TGTTTTCTTA	ATTGTCCGGG

1001 ATAAGGACGG CTACCGCACC GCCATGTCTT CCGAATTTGC CAAGCTGCTC  
1051 CGCTAA

This encodes a protein having amino acid sequence <SEQ ID 450>:

5           1 MPSETRLPNL IRALIFALGF IFLNACSEQT AQTVTLOGET MGTITYTVKYL  
          51 SNNRDKLPSP AKIQKRIDDA LKEVNRQMSY YQTDSEISRF NQHTAGKPLR  
101 ISSDFAHVTA EAVRLNRLTH GALDVTVGPL VNLWGFPGDK SVTREPSPEQ  
151 IKQAASVTGI DKIIQQQKGD YASLSKTHPK AYLDLSSIAK GFGVDKVGAGE  
201 LEKYGIQNYL VEIGGELHGK GKNAGHGPWR IGIEQPNIIQ GGNTQIIIVPL  
251 NNRLATSGD YRIFHVDKNG KRLSHIINPN NKRPIHNLA SISVSDSAM  
10       301 TADGLSTGLF VLGETEALRL AEQEKLAFL IVRDKDGYRT AMSSEFAKLL  
          351 R\*

This protein shows homology with a hypothetical lipoprotein precursor from *H. influenzae*:

15       sp|P44550|YOJL\_HAEIN HYPOTHETICAL LIPOPROTEIN HI0172 PRECURSOR >gi|1074292|pir|4  
          hypothetical protein HI0172 - Haemophilus influenzae (strain Rd KW20)  
          >gi|1573128 (U32702) hypothetical [Haemophilus influenzae] Length = 346  
          Score = 353 bits (896), Expect = 9e-97  
          Identities = 181/344 (52%), Positives = 247/344 (71%), Gaps = 4/344 (1%)  
  
20       Query: 7 LPNLIRALIFALGFIFLNACSEQTAQTVTLOGETMGTITYTVKYL SNNRDKLPSPAKIQKR 66  
          + LI +I + L AC ++T + ++L G+TMGTTY VKYL + S K +  
          Sbjct: 1 MKKLISGIIAVAMALSLAACQKET-KVISLSGKTMGTTYHVLYLDGSGITATSE-KTHEE 58  
  
25       Query: 67 IDDALKEVNRQMSYQTDSEISRFNQHT-AGKPLRISSDFAHVTA EAVRLNRLTHGALDV 125  
          I+ LK+VN +MSTY+ DSE+SRFNQ+T P+ IS+DFA V AEA+RLN++T GALDV  
          Sbjct: 59 IEAILKDVNAKMSYKKDSELSRFNQNTQVNTPIEISADFAKVLAEAIRLNKVTEGALDV 118  
  
30       Query: 126 TVGPLVNLWGFPGDKSVTREPSPEQIKQAASVTGIDKIIQQQKDYASLSKTHPKAYLDL 185  
          TVGP+VNLWGFPG+K ++P+PEQ+ + ++ GIDKI L K+ A+LSK P+ Y+DL  
          Sbjct: 119 TVGPPVNLWGFPGPEKRPEKQPTPEQLAERQAVGIDKITLDTNKEKATLSKALPQVYVDL 178  
  
35       Query: 186 SSIAGFGVDKVGAGELEKYGIQNYLVEIGGELHGKGKNAGHGPWRIGIEQPNIIQGGNTQ 245  
          SSIAGFGVD+VA +LE+ QNY+VEIGGE+ KGKN G+PW+I IE+P +  
          Sbjct: 179 SSIAGFGVDQVAEKLEQLNAQNYMVEIGGEIRAKGKNIEGKPPQIAIEKPTTGERAVE 238  
          Query: 246 IIVPLNNRLATSGDYRIFHVDKNGKRLSHIINPNKRPIHNLASISVSDSAMTADGL 305  
          ++ LNN +A+SGDYRI+ ++NGKR +H I+P PI H+LASI+V++ ++MTADGL  
          Sbjct: 239 AVIGLNNMGMASSGDYRIY-FEENGKRFHEIDPKTGYPHQHHLASITVLAPTSMTADGL 297  
  
40       Query: 306 STGLFVLGETEALRLAEQEKLAFLIVRDKDGYRTAMSSEFAKL 349  
          STGLFVLGE +AL +AE+ LAV+LI+R +G+ T SS F KL  
          Sbjct: 298 STGLFVLGEDKALEVAEKNLAVLYLIIRTDNGFVTKSSSAFCKL 341

Based on this analysis, it is predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 54

45 The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 451>:

50           1 ..CCGTGCCGCC GACAGGGCGA CGACGTGTAT GCGGCGCACG CGTCCCCTCA  
          51 AAAATTGTGG CTGCGCTTCA TCGGCGGCCG GTCGCATCAA AATATACGGG  
101 GCGGCGCGGC TCGGACGGG TGGCGCAAAG GCGTGCAAT CCGCGCGCAG  
151 GTGTTGTAC GGCAAAATGA AGGCAGCCKA yTGGCAATCG GCGTGATGGG  
201 CGGAGGGGCC GGCCAGCACG CwTCAGTCAA CGGCAAAGGC GGTGCGGCAG  
251 gCAGTGATT GTATGGTTAT GgCGGGGgTG TTTATGCTgC GTGGCATCAG  
301 TTGCGCGATA AACAAACGGG TgCGTATTTG GACGGCTGGT TGCAATACCA  
351 ACGTTTCAA CACCGCATCA ATGATGAAAA CCGTGCGGAA CgCTACAAAA  
401 CCAAAGGTTG GACGGCTTCT GTCGAAGGCG GCTACAACGC GCTTGTGGCG  
55       451 GAAGGCATTG TCGGAAAAGG CAATAATGTG CGGTTTTACC TACAACCGCA  
501 GgCGCAGTTT ACCTACTTGG GCGTAAACGG CGGCTTTACC GACAGCGAGG  
551 GGACGGCGGT CGGACTGCTC GGCAGCGGTC AGTGGCAAAG CCGCGCGGCG  
601 AtTCGGGCAA AAACCCGTTT TGCTTTGCGT AACGGTGTCA ATCTTCAGCC  
651 TTTTGCCGCT TTTAATGTTt TGCACAGGTC AAAATCTTTC GCGGTGGAAA  
60       701 TGGACGGCGA AAAACAGACG CTGGCAGGCA GGACGGCACT CGAAGGGCGG

751 TTCGGTATTG AAGCCGGTTG GAAAGGCCAT ATGTCCGCA..

This corresponds to the amino acid sequence <SEQ ID 452; ORF35>:

1 ..PCRRQGDDVY AAHASRQKLW LRFIGGRSHQ NIRGGAAADG WRKGVQIGGE  
5 51 VFVRQNEGSX LAIGVMGGRA GQHASVNGKG GAAGSDLYGY GGGVYAAWHQ  
101 LRDQKTGAYL DGWLQYQRFK HRINDENRAE RYKTKGWTAS VEGGYNALVA  
151 EGIVGKGNNV RFYLQPPAQF TYLGVNGGFT DSEGTAVGLL GSGQWQSRAG  
201 IRAKTRFALR NGVNLQPFPA FNVLHRSKSF GVEMDGEKQT LAGRTALEGR  
251 FGIEAGWKGH MSA..

Computer analysis of this amino acid sequence gave the following results:

10 Homology with putative secreted VirG-homologue of *N. meningitidis* (accession number A32247)

ORF and virg-h protein show 51% aa identity in 261aa overlap:

Orf35 5 QGDDVYAAHASRQKLWLRFIGGRSHQNIIRGGAA-ADGWRKGVQIGGEVFVRQNEGSXLAI 63  
+ D++ R+ LWLR I G S+Q ++G A +G+RKGVQ+GGEVF QNE + L+I  
15 virg-h 396 KNSDIFDRTLPRKGLWLRVIDGHSNQWVQKGTAPVEGYRKGVLGGEVFTWQNESNQLSI 455  
Orf35 64 GVMGGRAGQHASVNGKG--GAAGSDLYGYGGGVYAAWHQLRDQKTGAYLDGWLQYQRFKH 121  
G+MGG+A Q ++ + ++ G+G GVYA WHQL+DKQTGAY D W+QYQRF+H  
20 virg-h 456 GLMGGQAEQRSTFHNPDNDLTGNVKGFGAGVYATWHQLQDKQTGAYADSWMQYQFRH 515  
Orf35 122 RINDENRAERYKTKGWTASVEGGYNALVAEGIVGKGNNVRFYLQPPAQFTYLGVNGGFTD 181  
RIN E+ ER+ +KG TAS+E GYNAL+AE KGN++R YLQPPAQ TYLGVNG F+D  
25 virg-h 516 RINTEDGTERFTSKGITASIEAGYNALLAEHFTKKGNSLRVYLQPPAQLTYLGVNGKFSD 575  
Orf35 182 SEGTAVALLLGSGQWQSRAGIRAKTRFALRNGVNLQPFPAFNVLHRSKSGVEMDGEKQTL 241  
SE V LLGS Q Q+R G++AK +F+L + ++PFAA N L+ +K FGVEMDGE++ +  
30 virg-h 576 SENAHVNLGSRQLQTRVGQAKAQFSLYKNIAIEPFAAVNALYHNKPFVEMDGERRVI 635  
Orf35 242 AGRTALEGRFGIEAGWKGHMS 262  
+TA+E + G+ K H++  
30 virg-h 636 NNKTAIESQLGVAVKIKSHLT 656

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF35 shows 96.9% identity over a 259aa overlap with an ORF (ORF35a) from strain A of *N. meningitidis*:

35 orf35.pep 10 20 30  
PCRRQGDDVYAAHASRQKLWLRFIGGRSHQNIIRG  
:||||||| |||||||||||||||||||  
orf35a QRLAIPAEAVLYAQQAAYAANTLFGLRAADRGDDVYAADPSRQKLWLRFIGGRSHQNIIRG  
310 320 330 340 350 360  
40 orf35.pep 40 50 60 70 80 90  
GAAADGWRKGVQIGGEVFVRQNEGSXLAIIGVMGGRAGQHASVNGKGGGAAGSDLYGYGGGV  
||||| ||||||| ||||||| ||||||| ||||||| :|||||  
orf35a GAAADGRRKGVQIGGEVFVRQNEGSRLAIGVMGGRAGQHASVNGKGGGAAGSYLHGYGGGV  
45 370 380 390 400 410 420  
orf35.pep 100 110 120 130 140 150  
YAAWHQLRDQKTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAEGIV  
||||| ||||||| ||||||| ||||||| ||||||| :|||  
50 orf35a YAAWHQLRDQKTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAEGVV  
430 440 450 460 470 480  
orf35.pep 160 170 180 190 200 210  
GKGNNVRFYLQPPAQFTYLGVNGGFTDSEGTAVGLLGGSGQWQSRAGIRAKTRFALRNGVN  
||||| ||||||| ||||||| ||||||| ||||||| |||  
55 orf35a GKGNNVRFYLQPPAQFTYLGVNGGFTDSEGTAVGLLGGSGQWQSRAGIRAKTRFALRNGVN  
490 500 510 520 530 540  
orf35.pep 220 230 240 250 260  
LQPFPAFNVLHRSKSGVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSA  
60 ||||||| ||||||| ||||||| ||||||| |||||||

-272-

orf35a LQPFAAFNVLHRSKSFGEVMDGEKQTLAGRTALEGRFGIEAGWKGHMSARIGYGKRTDGD  
550 560 570 580 590 600

orf35a KEAALSLKWLFX  
610 620

The complete length ORF35a nucleotide sequence <SEQ ID 453> is:

1 ATGTTTCAGAG CTCAGCTTGG TTCAAATACT CGTCTACCA AAATCGGCGA  
51 CGATGCCGAT TTTTCATTTT CAGACAAGCC GAAACCCGGC ACTTCCCATT  
101 ATTTTTCAG CCGTAAAACC GATCAAAATT CATCCGAATA TGGGTATGAC  
151 GAAATCAATA TCCAAGGTAA AAATAACAAT AGCGGCATAC TCGCCGTCGA  
201 TAATATGCCC GTTGTTAAGA AATATATTAC AGATACTTAC GGGGATAATT  
251 TAAAGGATGC GGTTAAGAAG CAATTACAGG ATTTATACAA AACAAGACCC  
301 GAAGCTTGGG AAGAAAATAA AAAACGGACT GAGGAGGCGT ATATAGAACA  
351 GCTTGGACCA AAATTTAGTA TACTCAAACA GAAAAACCCG GATTTAATTA  
401 ATAAATTGGT AGAAGATTCC GTACTCACTC CTCATAGTAA TACATCACAG  
451 ACTAGTCTCA ACAACATCTT CAATAAAAAA TTACACGTCA AAATCGAAAA  
501 CAAATCCAC GTCGCCGAC AGGTGTTGGA ACTGACCAAG ATGACGCTGA  
551 AAGATTCCCT TTGGGAACCG CGCCGCCATT CCGACATCCA TATGCTGGAA  
601 ACTTCCGATA ATGCCCGCAT CCGCCTGAAC ACGAAAGATG AAAAAGTAC  
651 CGTCCATAAA GCGTATCAGG GCGGTGCGGA TTTCTGTTC GGCTACGACG  
701 TGCGGGAGTC GGACAAACCC GCCCTGACCT TTGAAGAAAA AGTCAGCGGA  
751 CAATCCGCG TGGTTTGGGA ACGCCGGCCG GAAAATCTGA AAACGCTCGA  
801 CGGGCGCAAA CTGATTGCGG CGGAAAAGGC AGACTCTAAT TCGTTTGCCT  
851 TTAAACAAAA TTACCGGCAG GACTGTACG AATTATTGCT CAAGCAATGC  
901 GAAGGCGGAT TTTGCTTGGG CGTGCAGCGT TTGGCTATCC CCGAGGCGGA  
951 AGCGGTTTTA TATGCCCAAC AGGCTTATGC GGCAATACT TTGTTGGGGC  
1001 TCGTGCCGC CGACAGGGC GACGACGTGT ATGCCGCCGA TCCGTCCCCT  
1051 CAAAAATTGT GGCTGCGCTT CATCGGCGGC CGGTGCGATC AAAATATACG  
1101 GGGCGGCGCG GCTGCGGACG GCGGCGCAA AGGCGTGCAA ATCGGCGGCG  
1151 AGGTGTTGT ACGGCAAAAT GAAGGCAGCC GGCTGGCAAT CCGCGTGATG  
1201 GGCGGCAGGG CTGGCCAGCA CGCATCAGTC AACGGCAAAG GCGGTGCGGC  
1251 AGGCAGTAT TTGCATGTT ATGGCGGGG TGTATTGCT GCGTGGCATC  
1301 AGTTGCGCGA TAAACAAACG GGTGCGTATT TGACCGCTG GTTGCAATAC  
1351 CAACGTTTCA AACACCGCAT CAATGATGAA AACCGTGCGG AACGCTACAA  
1401 AACCAAGGT TGGACGGCTT CTGTGCAAGG CGGCTACAAC GCGCTTGTGG  
1451 CGGAAGGCGT TGTCGGAAAA GGCAATAATG TGCGGTTTTA CCTGCAACCG  
1501 CAGGCGCAGT TTACCTACTT GGGCGTAAAC GCGGCTTTA CCGACAGCGA  
1551 GGGGACGCGG GTCGGACTGC TCGGACGCGG TCAGTGCAA AGCCGCGCCG  
1601 GCATTCGGGC AAAAACCCGT TTTGCTTTGC GTAACGGTGT CAATCTTCAG  
1651 CTTTTTCCG CTTTTAATGT TTTGCACAGG TCAAAATCTT TCGGCGTGGA  
1701 AATGGACGGC GAAAAACAGA CGCTGGCAGG CAGGACGGCG CTCGAAGGGC  
1751 GGTTCGGCAT TGAAGCCGGT TGGAAAGGCC ATATGTCCGC ACGCATCGGA  
1801 TACGGCAAAA GGACGGACGG CGACAAAGAA GCCGATTGT CGCTCAAATG  
1851 GCTGTTTTGA

This encodes a protein having amino acid sequence <SEQ ID 454>:

1 MFRAQLGSNT RSTKIGDDAD FSFSDKPKPG TSHYFSSGKT QNSSEYGYD  
51 EINIQQKNYN SGILAVDNMP VVKYITDTY GDNLKDAVKK QLQDLYKTRP  
101 EAWENKKRT EEAYIEQLGP KFSILKQKNP DLINKLVEDS VLTPHSNTSQ  
151 TSLNNIFNKK LHVKIENKSH VAGQVLELTK MTLKDSLWEP RRHSDIHMLE  
201 TSDNARIRLN TKDEKLTVHK AYQGGADFLF GYDVRESDKP ALTFEEKVSG  
251 QSGVVLERRP ENLKTLDGRK LIAAEKADSN SFAFKQNYRQ GLYELLKQK  
301 EGGFCLGVQR LAIPEAEAVL YAQQAYAANT LFLRAADRG DDVYAADPSR  
351 QKLWLRFIGG RSHQNIIRGA AADGRRKGVQ IGGEVFRQON EGSRLAIGVM  
401 GGRAGQHASV NGKGAAGSY LHGYGGGVYA AWHQLRDKQT GAYLDGWLQY  
451 QRFKHRINDE NRAERYKTKG WTASVEGGYN ALVAEGVVGK GNNVRFYLP  
501 QAQFTYLVN GGFTDSEGTA VLLGSGQWQ SRAGIRAKTR FALRNGVNLQ  
551 PFAAFNVLHR SKSFGVEMDG EKQTLAGRTA LEGRFGIEAG WKGHMSARIG  
601 YGKRTDGDKE AALSLKWLF\*

Homology with a predicted ORF from *N.gonorrhoeae*

ORF35 shows 51.7% identity over a 261aa overlap with a predicted ORF (ORF35ngn) from *N.gonorrhoeae*:

orf35.pep PCRRQGDDVYAAHASRQKLWLRFIGGRSHQNIIRG 34  
:::|:: |::|::| |::|::|  
orf35ngn FTKVQERDDIAIYAQQAQAANTLFALRLNDKNSDIFDRTLPRKGLWLRVIDGHSNQWVQG 370



5	orf35.pep	GAA-ADGWRKGQVIGGEVFTVRQNEGSXLAIGVMGGRAGQHASVNGKG--GAAGSDLYGYG	91
	orf35ngh	: :::     :    :   ::  : : ::   ::: : : :::  :	
	orf35.pep	GGVYAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAE	151
	orf35ngh	:    :    :    : : : : : : :    :  : :    :: : : : :	
10	orf35.pep	AGVYATWHQLQDKQTGAYVDSWMQYQRFRRHRINTEYATERFTSKGITASIEAGYNALLAE	490
	orf35ngh		
	orf35.pep	GIVGKGNVRFYLPQAQFTYLVNNGGFTDSEGTAVGLLGSQWQSRAGIRAKTRFALRN	211
	orf35ngh	::   :::   ::: :	
15	orf35.pep	HFTKKGNLSRVYLPQAQLTYLVNGKFSSENAQVNLLGSRQLQSRVGVQAKAQFAFTN	550
	orf35ngh		
	orf35.pep	GVNLQPFAAFNVLHRSKSGFVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSA	263
	orf35ngh	:::   :::   ::: : : : : : : : : : : : : : : : : : :	
	orf35.pep	GVTQFPFVAVNSIYQKPFGEVDGDRRVINNKTVIETQLGVAARIKSHLTLQASFNQRT	610
	orf35ngh		

A partial ORF35ngh nucleotide sequence <SEQ ID 455> is predicted to encode a protein having partial amino acid sequence <SEQ ID 456>:

20	1	..KKLRDRNSEY	WKEETYHIKS	NGRTYPNIPA	LFPKHPFDPF	ENINNSKKIS
	51	FYDKEYTEDY	LVGFARGFGV	EKRNGEEKP	LRQYFKDCVN	TENSNDNDCK
	101	ISSFGNYGPI	LIKSDIFALA	SQIKNSHINS	EILSVGNYIE	WLRPITLNKLT
	151	GWQEHLYAGL	DPFHYIEVTD	NSHVIGQITD	LGALELTNSL	WKPRWNSNID
25	201	YLITKNAEIR	FTNKESLLV	KEDYAGGARF	RFAYDLKDKV	PEIPVLTFEKE
	251	NITGTSDIIF	EGKALDNLKH	LDGHQIVKVN	DTADKDAFRL	SSKYRKGIYT
	301	LSLQQRPEGF	FTKVQERDDI	AIYAQAQAAQ	NTLFALRLND	KNSDIFDRTL
	351	PRKGLWLRLV	DGHSNQWVQG	KTAPVEGYRK	GVQLGGEVFT	WQNESNQLSI
30	401	GLMGGQAEQR	STFRNPDTDN	LTTGNVKGFG	AGVYATVHQL	LQDKGTGAYVD
	451	SWMQYQFRFH	RINTEYATER	FTSKGITASI	EAGYNALLAE	HFTKKGNSLR
	501	VYLQPPAQLT	YLGVNGKFSD	SENAQVNLLG	SRQLQSRVGV	QAKAQFAFTN
	551	GVTFPFVAV	NSIYQKQKPF	VEIDGDRRVI	NNKTVIETQL	GVAAKIKSHL
	601	TLOASENROT	SKHHHAKOGA	LNLOWTF*		

Based on this prediction, these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 55

35 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 457>:

40

1	..GCGGAATATG	TTCAGTTC	TATAGATTG	TTCAGTGTG	GTAAATCGG
51	GGGCGGTATA	CCTAAGCTA	AGCCTGTGT	TGATGCGAAA	CCGAGATGGG
101	AGGTTGATAG	GAAGCTTAAT	AAATTGACAA	CTCGTGAGCA	GGTGGAGAAA
151	AATGTTACAG	AAACGAGAAG	AAGGAGTCAG	AGTAGTCAGT	TAAAGCCCCA
201	TGCGCAACGA	GAATGGGAAA	ATAAAACAGG	GTTAGATTTT	AATCATTTTA
251	TAGGTGGTGA	TATCAATAAA	AAAGGCACAG	TAACAGGAGG	GCATAGTCTA
301	ACCCGTGGTG	ATGTACGGGT	GATACAACAA	ACCTCGGCAC	CTGATAAAÇA
351	TGGGGT.TTA	TCAAGCGACA	GTGGAAATTN	A	

This corresponds to the amino acid sequence <SEQ ID 458; ORF46>:

45           1    ..AEYVQFSIDL FSVGKSGGGI PKAKPVFDAK PRWEVDRKLN KLTTRQEVEK  
          51    NVQETRRRSQ SSQFKAHAQR EWENKTGLDF NHFIGGDINK KGTVTGGHSL  
         101    TRGDVVRVIO TSAPDKHGXL SSDSGNX

Further work revealed further partial nucleotide sequence <SEQ ID 459>:

50	1	..GCAGTGTGCC	TnCCGATGCA	TGCACACGCC	TCAnATTTGG	CAAAcGATTc
	51	TTTTATCCGG	CAGGTTCTCG	ACCGTCAGCA	TTTCGAACCC	GACGGGAAAT
	101	ACCACCTATT	CGGCAGCAGG	GGGGAACCTG	CCGAGCGCTCA	GTCTCATATC
	151	GGATTGGGAA	AAATACAAAG	CCATCAGTTG	GGCAACCTGA	TGATTCAACA
55	201	GGCGGCCATT	AAAGGAAATA	TCGGCTACAT	TGTCGGCTTT	TCCGATCACG
	251	GGCACGAAGT	CCATTCcCCc	TTcGACAAcC	ATGCCTCACA	TTCCGATTCT
	301	GATGAAGCCG	GTAgtCCCGT	TGACGGATTT	AGcCTTTACc	GCATCCATTG
	351	GGACGGATAC	GAACACCATC	CCGCCGACGG	CTATGACGGG	CCACAGGGCG
	401	CGCGGTATCC	CGCTCCCAAA	GGCGCGAGGG	ATATATACAG	TTACGACATA

-274-

5  
 451 AAAGGCGTTG CCCAAAATAT CCGCCTCAAC CTGACCGACA ACCGCAGCAC  
 501 CGGACAACGG CTTGCCGACC GTTCCACAA TGCCGGTAGT ATGCTGACGC  
 551 AAGGAGTAGG CGACGGATTC AAACGCGCCA CCCGATACAG CCCCAGCTG  
 601 GACAGATCGG GCAATGCCGC CGAAGCCTTC AACGGCACTG CAGATATCGT  
 651 TAAAAACATC ATCGGCGCTG CAGGAGAAAT TGT

This corresponds to the amino acid sequence <SEQ ID 460; ORF46-1>:

10  
 1 ..AVCLPMHAHA SXLANDSFIR QVLDRQHFEP DGKYHLFGSR GELAERQSHI  
 51 GLGKIQSHQL GNLMIQAAI KGNIGYIVRF SDHGHEVHSP FDNHASHSDS  
 101 DEAGSPVDGF SLYRIHWDGY EHHFADGYDG PQGGGYPAK GARDIYSYDI  
 151 KGVAQNIRLN LTNRSTGQR LADRFHNAGS MLTQGVGDGF KRATRYSPLE  
 201 DRSGNAAEAF NGTADIVKNI IGAAGEI

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF46 shows 98.2% identity over a 111aa overlap with a predicted ORF (ORF46ng) from *N.*

15 *gonorrhoeae*:

20  
 orf46.pep AEYVQFSIDLFSVGKSGGGIPKAKPVFDAKPRWEVDRLNKLTTTR 45  
 orf46ng PKTGVFPDGGKGFNFEEKHVKYDTKLDIQELSGGGIPKAKPVFDAKPRWEVDRLNKLTTTR 217  
 20 orf46.pep EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGTVTGGHSLTRGDV 105  
 orf46ng EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGAVTGGHSLTRGDV 277  
 25 orf46.pep RVIQQTSAAPDKHGXLSDDSGN 126  
 orf46ng RVIQQTSAAPDKHGVLSDDSGN 298

A partial ORF46ng nucleotide sequence <SEQ ID 461> is predicted to encode a protein having partial amino acid sequence <SEQ ID 462>:

30  
 1 ..RRLKHCCHAR LGSFAHRKQD GAHQRFGRYG ATQRLCRSSH PRLGSPKPCQ  
 51 RTRHRSRQQY LYGSHPHQRD WSCPGKIQLG RHGHTSCRAV ADXRDRICER  
 101 EIRRQRQXCR CRLGKIPSL S IPKYPLKLEQ RYKGENITSS TVPPSNGKNV  
 151 KLADQRHPKT GVPFDGKGF NFEKHKVYDT KLDIQELSGG GIPKAKPVFD  
 201 AKPRWEVDK LNKLTTRQV EKNVQETRRR SQSSQFKAHA QREWENKTGL  
 251 DFNHFIGGDI NKKGAVTGGH SLTRGDV RVI QQTSAAPDKHGXLSDDSGN\*

35 Further work revealed the complete gonococcal DNA sequence <SEQ ID 463>:

40  
 1 TTGGGCATTT CCCGAAAAT ATCCCTTATT CTGTCCATAC TGGCAGTGTG  
 51 CCTGCCGATG CATGCACACG CCTCAGATTT GGcaAACGAT CCCTTTATCC  
 101 GgCaggttcT CGaccGTCAG CATTTCGaac ccgacggGAa ATACCaCCTA  
 151 TTcggCaGCA GGGGGGAGCT TgccnagcGC aacggccATa tcggattggG  
 201 aaacaTAcAa Agccatcagt tGggccacct gatgattcaa caggcggccg  
 251 ttgaaggaaA TAtcgGctac attgtccgct tttccgatca cgggcacaaa  
 301 ttccattcgc ccttcGAcAa ccaTGCTCA CATTCCGATT CTGACGAAGC  
 351 CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CCGCATCCAT TGGGACGGAT  
 401 ACGAACACCA TCCCGCCGAC GGCTATGACG GGCCACAGGG CGGCGGCTAT  
 45 451 CCCGCTCCCA AAGGCGCGAG GGATATATAC AGCTACGACA TAAAAGGCGT  
 501 TGCCCAAAAT ATCCGCTCA ACCTGACCGA CAACCGCAGC ACCGGACAAC  
 551 GGCTTGCCGA CCGTTTCAC AATGCCGCG CTATGCTGAC GCAAGGAGTA  
 601 GGCGACGGAT TCAAACGCGC CACCCGATAC AGCCCCGAGC TGGACAGATC  
 651 GGGCAATGCc gccGAAGCCT TCAACGCGAC TGCAGATATC GTCAAAAACA  
 701 TCATCGGCGC GGCAGGAGAA ATTGTCCGCG CAGGCGATGC CGTGCagGGT  
 751 ATAAGCGAAG GCTCAAACAT TGCTGTCATG CACGGCTTGG GTCTGCTTTC  
 801 CACCGAAAAC AAGATGGCGC GCATCAACGA TTTGGCAGAT ATGGCGCAAC  
 851 TCAAAGACTA TGCCGAGCA GCCATCCGCG ATTGGGCAGT CCAAAACCCC  
 901 AATGCCGCAC AAGGCATAGA AGCCGTCAGC AATATCTTTA TGGCAGCCAT  
 55 951 CCCCATCAAA GGGATTGGAG CTGTCCGGGG AAAATACGGC TTGGGCGGCA  
 1001 TCACGGCACA TCCTGTCAAG CGGTCGAGA TGGGCGCGAT CGCATTGCCG  
 1051 AAAGGGAAT CCGCCGTCAG CGACAATTTT GCCGATGCGG CATACGCCAA  
 1101 ATACCCGTCC CCTTACCATT CCCGAAATAT CCGTTCAAAC TTGGAGCAGC

-275-

1151 GTTACGGCAA AGAAAACATC ACCTCCTCAA CCGTGCCGCC GTCAAACGGC  
 1201 AAAAATGTCA AACTGGCAGA CCAACGCCAC CCGAAGACAG GCGTACCGTT  
 1251 TGACGGTAAA GGGTTTCCGA ATTTTGAGAA GCACGTGAAA TATGATACGA  
 1301 AGCTCGATAT TCAAGAATTA TCGGGGGGCG GTATACCTAA GGCTAAGCCT  
 5 1351 GTGTTTGATG CGAAACCGAG ATGGGAGGTT GATAGGAAGC TTAATAAATT  
 1401 GACAACTCGT GAGCAGGTGG AGAAAAATGT TCAGGAAACG AGAAGAAGGA  
 1451 GTCAGAGTAG TCAGTTTAAA GCCCATGCGC AACGAGAATG GGAAAAATAA  
 1501 ACAGGGTTAG ATTTTAATCA TTTTATAGGT GGTGATATCA ATAAGAAAGG  
 1551 CACAGTAACA GGAGGGCATA GTCTAACCCG TGGTGATGTA CGGGTGATAC  
 10 1601 AACAAACCTC GGCACCTGAT AAACATGGGG TTTATCAAGC GACAGTGGA  
 1651 ATTAATAAGC CTGATGGAAG TTGGGAGGTG AAAACGAAAA AAGGTGGGAA  
 1701 AGTGATGACC AAGCACACCA TGTTCCCAA AGATTGGGAT GAGGCTAGAA  
 1751 TTAGGGCTGA AGTTACTTCG GCTTGGGAAA GTAGAATAAT GCTTAAGGAT  
 1801 AATAAATGGC AGGGTACAAG TAAATCGGGT ATTAATAATAG AAGGATTTAC  
 15 1851 CGAACCTAAT AGAACAGCAT ATCCCATTTA TGAATAG

This corresponds to the amino acid sequence <SEQ ID 464; ORF46ng-1>:

1 LGISRKISLI LSILAVCLPM HAHASDLAND PFIRQVLDRO HFEPDGKYHL  
 51 FGSRGELAXR NGHIGLGNIQ SHQLGHLMIQ QAAVEGNIGY IVRFSHDHGHK  
 101 FHSPFDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHHPAD GYDGPQGGGY  
 20 151 PAPKGARDIY SYDIKVAQN IRLNLTNRS TGQRLADRFH NAGAMLTQGV  
 201 GDGFKRATRY SPELDRSGNA AEFNGTADI VKNIIGAAGE IVGAGDAVQG  
 251 ISEGSNIAVM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP  
 301 NAAQIEAVS NIFMAAPIK GIGAVRGKYG LGGITAHVPK RSQMGAIALP  
 351 KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYGKENI TSSTVPPSNG  
 25 401 KNVKLADQRH PKTGVPPFDGK GFFNFEKHVK YDTKLDIQEL SGGGIPKAKP  
 451 VFDAPRWEV DRKLNKLTR EQVEKNVQET RRRSQSSQFK AHAQREWENK  
 501 TGLDFNHFIF GDINKKGTVT GGHSLTRGDV RVIQOTSAPD KHGVYQATVE  
 551 IKKPDGSWEV KTKKGGKVM THTMFPKDW EARIARAEVTS AWESRIMLKD  
 601 NKWQGTSKSG IKIEGFTEPN RTAYPIYE\*

30 ORF46ng-1 and ORF46-1 show 94.7% identity in 227 aa overlap:

		10	20	30	40
orf46-1.pep		AVCLPMHAHASXLANDSFIRQVLDROHFEPDGKYHLFGSRGELAER			
orf46ng-1	LGISRKISLILSILAVCLPMHAHASDLANDPFIRQVLDROHFEPDGKYHLFGSRGELAXR	10	20	30	40
35		50	60	70	80
orf46-1.pep	QSHIGLGKIQSHQLGNLMIQQAIAKGNIGYIVRFSHDHGHEVHSPFDNHASHSDSDEAGSP	50	60	70	80
40	orf46ng-1	NGHIGLGNIQSHQLGHLMIQQAIVEGNIGYIVRFSHDHGKHFHSPFDNHASHSDSDEAGSP	90	100	110
		110	120	130	140
orf46-1.pep	VDGFSLYRIHWDGVEHHHPADGYDGPQGGGYPAPKGARDIYSYDIKVAQNIRLNLTNRS	110	120	130	140
45	orf46ng-1	VDGFSLYRIHWDGVEHHHPADGYDGPQGGGYPAPKGARDIYSYDIKVAQNIRLNLTNRS	150	160	170
		170	180	190	200
orf46-1.pep	TGQRLADRFHNAAGMLTQGVGDGFKRATRYSPELDRSGNAAEFNGTADIVKNIIGAAGE	170	180	190	200
50	orf46ng-1	TGQRLADRFHNAAGMLTQGVGDGFKRATRYSPELDRSGNAAEFNGTADIVKNIIGAAGE	210	220	230
		250	260	270	280
orf46-1.pep	I	250	260	270	280
55	orf46ng-1	IVGAGDAVQGISSEGSNIAVMHGLGLLSTENKMARINDLADMAQLKDYAAAIRDWAVQNP	290	300	
60		250	260	270	280

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF46ng-1 shows 87.4% identity over a 486aa overlap with an ORF (ORF46a) from strain A of *N. meningitidis*:

10 20 30 40 50 60

-276-

	orf46a.pep	LGISRKISLILSILAVCLPMHAHASDLANDSFIRQVLDROHFEPDGKYHLFGSRGELAER
	orf46ng-1	LGISRKISLILSILAVCLPMHAHASDLANDPFIRQVLDROHFEPDGKYHLFGSRGELAXR
5		10 20 30 40 50 60
	orf46a.pep	70 80 90 100 110 120
	orf46ng-1	SGHIGLGNIQSHQLGNLFIQAAIKGNIGYIVRFSHDHGHEVHSPFDNHASHSDSDEAGSP
10		:     :     :     :     :     :     :     :     :     :
	orf46ng-1	NGHIGLGNIQSHQLGHLMIQAAVEGNIGYIVRFSHDHGKHFHSPFDNHASHSDSDEAGSP
		70 80 90 100 110 120
	orf46a.pep	130 140 150 160 170 180
15		VDGFSLYRIHWDGYEHHPADGYDGPQGGYPAPKGARDIYSYDIKGVAQNIRLNLTDNRS
	orf46ng-1	VDGFSLYRIHWDGYEHHPADGYDGPQGGYPAPKGARDIYSYDIKGVAQNIRLNLTDNRS
		130 140 150 160 170 180
	orf46a.pep	190 200 210 220 230 240
20		TGQRLVDRFHNHTGSMLTQGVGDGFKRATRYSPELDMSGNAEAFNGTADIVKNIIGAAGE
	orf46ng-1	TGQRLADRFHNAGAMLTQGVGDGFKRATRYSPELDMSGNAEAFNGTADIVKNIIGAAGE
		190 200 210 220 230 240
	orf46a.pep	250 260 270 280 290 300
25		IVGAGDAVQGISSEGSNIAMVHGLGLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP
	orf46ng-1	IVGAGDAVQGISSEGSNIAMVHGLGLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP
		250 260 270 280 290 300
	orf46a.pep	310 320 330 340 350 360
30		NAAQGIEAVSNIFTAVIPVKIGIAVRGKYGLGGITAHVPKRSQMGEIALPKGKSAVSDNF
	orf46ng-1	NAAQGIEAVSNIFMAAIPKIGIAVRGKYGLGGITAHVPKRSQMGAIALPKGKSAVSDNF
		310 320 330 340 350 360
	orf46a.pep	370 380 390 400 410 420
35		ADAAYAKYPSPYHSRNIIRSNLEQRYGKENITSSTVPPSNGKNVKLANRHPKTKVPFDGK
	orf46ng-1	ADAAYAKYPSPYHSRNIIRSNLEQRYGKENITSSTVPPSNGKNVKLADQRHPKTKGVFPDGK
40		370 380 390 400 410 420
	orf46a.pep	430 440 450 460 470
45		GFPNFEKDVKYDTRINTAVPQVN----PIDEPVEN--PKGSVGSASHWSITARIQYAKLP
	orf46ng-1	GFPNFEKHVKYDTKLD--IQELSGGGIPKAKPVFDAKPRWEVDRKLN-KLTREQVEKNV
		430 440 450 460 470
	orf46a.pep	480 490 500 510 520 530
50		RQGRIRYIPPKNYSPSAPLPKGPNNGYLDKFGNEWTKGPSRTKGQEFWDVQLSKTGREQ
	orf46ng-1	QETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGTVTGGHSLTRGDVRIQOTS
		480 490 500 510 520 530

The complete length ORF46a DNA sequence <SEQ ID 465> is:

55	1	TTGGGCATT	CCCCGAAAT	ATCCCTTATT	CTGTCCATAC	TGGCAGTGTG
	51	CCTGCCGATG	CATGCACACG	CCTCAGATTT	GGCAAACGAT	TCTTTTATCC
	101	GGCAGGTTCT	CGACCGTCAG	CATTTTGAAC	CCGACGGGAA	ATACCACCTA
	151	TTCGGCAGCA	GGGGGGAAC	TGCCGAGCGC	AGCGGTCATA	TCCGATTGGG
	201	AAACATACAA	AGCCATCAGT	TGGGCAACCT	GTTCATCCAG	CAGGCGGCCA
	251	TTAAAGGAAA	TATCGGCTAC	ATTGTCCGCT	TTTCCGATCA	CGGCACGAA
60	301	GTCCATTCCC	CCTTCGACAA	CCATGCCTCA	CATTCCGATT	CTGATGAAGC
	351	CGGTAGTCCC	GTTGACGGAT	TCAGCCTTTA	CCGCATCCAT	TGGGACGGAT
	401	ACGAACACCA	TCCCGCCGAC	GGCTATGACG	GGCCACAGGG	CGGCGGCTAT
	451	CCCCTCCCA	AAGGCGCGAG	GGATATATAC	AGCTACGACA	TAAAAGGCGT
	501	TGCCCAAAAT	ATCCGCTCA	ACCTGACCGA	CAACCGCAGC	ACCGGACAAC
65	551	GGCTTGTCTGA	CCGTTTCCAC	AATACCGGTA	GTATGCTGAC	GCAAGGAGTA
	601	GGCGACGGAT	TCAAACGCGC	CACCCGATAC	AGCCCCGAGC	TGGACAGATC
	651	GGGCAATGCC	GCCGAAGCTT	TCAACGGCAC	TGCAGATATC	GTCAAAAACA
	701	TCATCGGCGC	GGCAGGAGAA	ATTGTCGGCG	CAGGCGATGC	CGTGCAAGGT
	751	ATAAGCGAAG	GCTCAAACAT	TGCTGTTATG	CACGGCTTGG	GTCTGCTTTC
70	801	CACCGAAAAC	AAGATGGCGC	GCATCAACGA	TTTGGCAGAT	ATGGCGCAAC

-277-

851 TC AAGACTA TGCCGCAGCA GCCATCCGCG ATTGGGCAGT CCAAAACCCC  
 901 AATGCCGCAC AAGGCATAGA AGCCGTCAGC AATATCTTTA CGGCAGTCAT  
 951 CCCCCTCAA GGGATTGGAG CTGTTCCGGG AAAATACGGC TTGGGCGGCA  
 1001 TCACGGCACA TCCTGTCAAG CGGTCGCAGA TGGGCGAGAT CGCATTGCCG  
 5 1051 AAAGGGAAAT CCGCCGTCAG CGACAATTTT GCCGATGCGG CATACGCCAA  
 1101 ATACCCGTCC CTTACCAAT CCCGAAATAT CCGTTCAAAC TTGGAGCAGC  
 1151 GTTACGGCAA AGAAAACATC ACCTCCTCAA CCGTGCCGCC GTCAAACGGA  
 1201 AAGAATGTGA AACTGGCAA CAAACGCCAC CCGAAGACCA AAGTGCCGTT  
 1251 TGACGGTAAA GGGTTTCCGA ATTTTGAAAA AGACGTAAAA TACGATACGA  
 10 1301 GAATTAATAC CGCTGTACCA CAAGTGAATC CTATAGATGA ACCCGTCTTT  
 1351 AATCCTAAAG GTTCTGTCGG ATCGGCTCAT TCTTGGTCTA TAAGTGCCAG  
 1401 AATTCAATAC GCAAATATC CAAGGCAAGG TAGAATCAGA TATATCCCAC  
 1451 CTAAAAATTA CTCTCCTTCA GCACCGCTAC CAAAAGGACC TAATAATGGA  
 1501 TATTTGGATA AATTTGGTAA TGAATGGACT AAAGGTCCAT CAAGAACTAA  
 15 1551 AGGTCAAGAA TTTGAATGGG ATGTTCAATT GTCTAAACA GGAAGAGAGC  
 1601 AACTTGGATG GGCTAGTAGG GATGGTAAGC ATTTAAATAT ATCAATTGAT  
 1651 GGAAAGATTA CACACAAATG A

This corresponds to the amino acid sequence <SEQ ID 466>:

20 1 LGISRKISLI LSILAVCLPM HAHASDLAND SFIRQVLDRO HFEPDGKYHL  
 51 FGSRGELAER SGHIGLGNIO SHQLGNLFIO QAAIKGNIGY IVRFSDHGHE  
 101 VHSPFDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHHPAD GYDGPQGGGY  
 151 PAPKGARDIY SYDIKGVAQN IRLNLTNRS TGQRLVDRFH NTGSMLTQGV  
 201 GDGFKRATRY SPELDRSNA AEAFTGTADI VKNIIGAAGE IVGAGDAVQG  
 25 251 ISEGSNIAVM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP  
 301 NAAQIEAVS NIETAVIPVK GIGAVRGKYG LGGITAHVPK RSQMGEIALP  
 351 KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYGKENI TSSTVPPSNG  
 401 KNVKLANKRH PKTKVPFDGK GFPNFEKDVK YDTRINTAVP QVNPIDEPVF  
 451 NPKGSVGSAA SWSITARIQY AKLPRQGRIR YIPPKNYSPS APLPKGPNNG  
 501 YLDKFGNEWT KGPSRTKGQE FEWDVQLSKT GREQLGWASR DGKHLNISID  
 30 551 GKITHK\*

Based on this analysis, including the presence of a RGD sequence in the gonococcal protein, typical of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 35 Example 56

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 467>:

40 1 ATGAATATTC ACACCCTGCT CTCAAACAA TGGACGCTGC CGCCATTCCCT  
 51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTTGCC CCCAATGCGG  
 101 TGTTTTGGGT TTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT  
 151 TTGGACTATC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GGCGTTTCGT  
 201 CAAAATTGCC GCGGTATTGG CGTTTTGGCT GGCGGTTTTG TTTGACGGGC  
 251 TGATGATGGT GATCCAATC TTCCCTTTTA TGGATCTCAT CGGCGCCATC  
 301 AACCTCGTCC CCTTCATCCT GACCGCCCCC GCCCCTTATC AGATAATGAC  
 351 CGGGCTG...

45 This corresponds to the amino acid sequence <SEQ ID 468; ORF48>:

1 MNIHTLLSKQ WTLPPFLPKR LLSLLILLA PNAFWVLAL LTATARPIVN  
 51 LDYLPALLI ALPWRFKIA GVLAFLWAVL FDGLMMVIQL FPFMDLIGAI  
 101 NLVPFILTAP APYQIMTGL...

Further work revealed the complete nucleotide sequence <SEQ ID 469>:

50 1 ATGAATATTC ACACCCTGCT CTCAAACAA TGGACGCTGC CGCCATTCCCT  
 51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTTGCC CCCAATGCGG  
 101 TGTTTTGGGT TTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT  
 151 TTGGACTATC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GGCGTTTCGT  
 201 CAAAATTGCC GCGGTATTGG CGTTTTGGCT GGCGGTTTTG TTTGACGGGC  
 55 251 TGATGATGGT GATCCAATC TTCCCTTTTA TGGATCTCAT CGGCGCCATC  
 301 AACCTCGTCC CCTTCATCCT GACCGCCCCC GCCCCTTATC AGATAATGAC  
 351 CGGGCTGTTG CTGCTGTATA TGCTGGCGAT GCCGTTTGTG TTGCAGAAAG

-278-

5  
10  
15  
20

```

401 CCGCCGCCAA AACCGACTTC CGGCACATTG CCGTCTGCGC CGCCGTGTGTG
451 GCGGCAGCCG GCTATTTTAC CGGCCATTG AGTTACTACG ACCGGGGTCG
501 GATGGCCAAT ATCTTCGCGC CAAACAACCTT CTACTACGCC AAAAGTCAGG
551 CGATGCTCTA CACCGTCAGC CAGAATGCCG ACTTTATTAC CGCCGGCCTG
601 GTCGATCCCG TCTTCTCTCC CTTGGGCAAT CAACAGCGTG CCGCCACGCA
651 TCTGAACGAG CCGAAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT
701 GGGGGCTGCC GGCCAATCCC GAACTTCAAA ACGCCACTTT TGCCAAACTG
751 CTGGCGCAAA AAGACCGTTT TTCGGTTTGG GAAAGCGGCA GTTTTCCCTT
801 CATCGGCGCG ACGGTCGAAG GCGAAATGCG CGAACTGTGT GCCTACGGCG
851 GTTTGCGCGG GTTCGCACTG CGCGCGCGC CCGACGAAAA ATTTGCCCGC
901 TGCCTCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACCT TTGCGATGCA
951 CGGCGCGGGC AGTTCGCTTT ACGACCGCTT CAGCTGGTAT CCGAGGGCGG
1001 GCTTTCAAGA AATCAAAACC GCCGAAAACC TGATCGGTAA AAAAACCTGC
1051 GCCATTTTCG GCGGCGTGTG CGACAGCGAG CTGTTCCGCG AAGTGTCCGG
1101 ATTTTTTCAA AAACACGACA AGGGACTGTT TTACTGGATG ACGCTGACCA
1151 GCCACGCCGA CTATCCCGAA TCCGACATTT TCAACCACAG GCTCAAATGC
1201 ACCGAATATG GCCTGCCCCG CGAAACCGAC CTCTGCCGCA ATTTCAGCCT
1251 GCACACCCAA TTCTTCGACC AACTGGCGGA TTGATCCAA CGCCCCGAAA
1301 TGAAAGGCAC GGAAGTCATC ATCGTCGGCG ACCATCCGCC GCCCGTCGGC
1351 AACCTCAATG AAACCTTCCG CTACCTCAAA CAGGGGCACG TCGCCTGGCT
1401 GAACTTCAAA ATCAAATAA

```

This corresponds to the amino acid sequence <SEQ ID 470; ORF48-1>:

25  
30

```

1 MNIHTLLSKQ WTLPPFLPKR LLLSLILL PNAVFWLAL LTATARPIVN
51 LDYLPAAALI ALPWRVFKIA GVLAFWLAVL FDGLMMVIQL FPFMDLIGAI
101 NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAKTD F RHIAVCAAVV
151 AAAGYFTGHL SYDRGRMAN IFGANNFYA KSQAMLYTVS QNADFITAGL
201 VDPVFLPLGN QORAATHLNE PKSQILFIV AESWGLPANP ELQATFAKL
251 LAQKDRFSVW ESGSFPIGA TVEGEMRELC AYGLRGFAL RRAPDEKFA
301 CLPNRLQEG YATFAMHGAG SSLYDRFSWY PRAGFQEIKT AENLIGKTC
351 AIFGGVCDSE LFGEVSAFFK KHDKGLFYWM TLTSHADYPE SDIFNHLK
401 TEYGLPAETD LCRNFSLHTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPPVG
451 NLNETFRYLK QGHVWLNFK IK*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

35 ORF48 shows 94.1% identity over a 119aa overlap with an ORF (ORF48a) from strain A of *N. meningitidis*:

40  
45  
50

```

          10      20      30      40      50      60
orf48.pep MNIHTLLSKQWTLPPFLPKRLLLSLILLAPNAVFWVLALLTATARPIVNLDYLPAAALLI
          10      20      30      40      50      60
orf48a    MNIHTLLSKQWTLPPFLPKRLLLSLILLXPNVFWVLALLTATARPIVNLXYLPAAALLI

          70      80      90      100     110     119
orf48.pep ALPWRVFKIAGVLAFWLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGL
          70      80      90      100     110     120
orf48a    ALPWRXVKIXGLAXWLAVLFDGLMMVIQLFPFMDLIGAINLVPFIXTAPALYQIMTGLL

          130     140     150     160     170     180
orf48a    LLYMLAMPFVLQKAAAKTDFRHIAACA AVVVAAGYFTGHLSXYDRGRMANIFGANNFYA

```

The complete length ORF48a nucleotide sequence <SEQ ID 471> is:

55  
60

```

1 ATGAATATTC ACACCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCTT
51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTNNCC CCCAATGCGG
101 TGTTTTTGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
151 TTGGANTACC TTCCCGCGCG GCTGCTGATC GCCCTGCCTT GGCGTNTCGT
201 CAAAATTGNC GCGGTATTGG CGTNTTGGCT GCGGTTTTTG TTTGACGGGC
251 TGATGATGGT GATCCAATC TTCCCTTTTA TGGATCTCAT CGGCGGCATC
301 AACCTCGTCC CTTTCATCNT GACCGCCCCC GCCCTTTATC AGATAATGAC
351 CGGGCTGTTA CTGCTGTATA TGCTGGCGAT GCCGTTTGTG TTGCAGAAAG
401 CCGCCGCCAA AACCGACTTC CGACACATTG CCGCTGTGCG CGCCGTTGTG
451 GTGGCAGCCG GCTATTTTAC CGGCCATTG AGTTANTACG ACCGGGGGGC

```

-279-

501 GATGGCCAAT ATCTTCGGCG CAAACAATT CTATTACGCC AAAAGTCAGG  
 551 CGATGCTCTA CACCGTCAGC CAGAATGCCG ACTTTATTAC CGCCGGCCTG  
 601 GTCGATCCCG TCTTCTCCCG CTTGGGCAAT CAACAGCGTG CCGCCACGCA  
 651 TCTGAACGAG CCGAAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT  
 701 GGGGGCTGCC GGCCAATCCC GAACTTCAAA ACGCCACTTT TGCCAAACTG  
 751 CTGGCGCAAA AAGANCGTTT TTCGGTTTGG GAAAGCGGCA GTTTTCCCTT  
 801 CATCGGCGCG ACGATCGAAG GCGAAATGCG CGAACTGTGT GCCTACGGCG  
 851 GTTTGCGCGG GTTCGCACTG CGCCGCGCGC CCGACGAAAA ATTTGCCCGC  
 901 TGCCTCCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACCT TTGCGATGCA  
 951 CGGCGCGGGC AGTTCGCTTT ACGACCGCTT CAGCTGGTAT CCGAGGGCGG  
 1001 GCTTTCAAGA AATCAAAACC GCCGAAAACC TGATCGGTAA AAAAACCTGC  
 1051 GCCATTTTCG GCGGCGTGTG CGACAGCGAG CTGTTCTGGCG AAGTGTCTGGC  
 1101 ANTTTTTCAA AAACACGACA AGGGACTGTT TTAAGGATG ACGCTGACCA  
 1151 GCCACGCCGA CTATCCCGAA TCNGACATTT TCAACCACAG GCTCAAATGC  
 1201 ACCGAATATG GCCTGCCCGC CGAAACCGAC NTCTGCCGCA ATTTAGCCT  
 1251 GCACACCAA TTTCTCGACC AACTGGCGGA TTGATCCAA CGCCCCGAAA  
 1301 TGAAAGGCAC GGAAGTCATC ATCGTCGGCG ACCATCCGCC GCCCGTCTGGC  
 1351 AACCTCAATG AAACCTTCCG CTACCTCAAA CAGGGGCACG TCGNCTGGCT  
 1401 GAACTTCAAA ATCAATAA

20 This encodes a protein having amino acid sequence <SEQ ID 472>:

1 MNIHTLLSKQ WTLPPFLPKR LLLSLILLX PNAVFWLAL LTATARPIVN  
 51 LXLYPAALLI ALPWRXVKIX GVLAXWLAVL FDGLMMVIQL FPFMDLIGAI  
 101 NLVPFIXTAP ALYQIMTGLL LLYMLAMPFV LQKAAAKTDF RHIAACAADV  
 151 VAAGYFTGHL SXYDRGRMAN IFGANNFYYA KSQAMLYTVS QNADFITAGL  
 201 VDPVFLPLGN QQRAATHLNE PKSQKILFIV AESWGLPANP ELQNATFAKL  
 251 LAQKXRFVSW ESGSFPIGA TIEGEMRELC AYGGLRGFAL RRAPDEKFA  
 301 CLPNRLKQEG YATFAMHGAG SLYDRFSWY PRAGFQEIKT AENLIGKKT  
 351 AIFGGVCDSE LFGEVSAXFK KHDKGLFYWM TLTSHADYPE SDIFNHLK  
 401 TEYGLPAETD XCRNFSLHTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPPVG  
 451 NLNETFRYLK QGHVXWLNFK IK\*

ORF48a and ORF48-1 show 96.8% identity in 472 aa overlap:

	10	20	30	40	50	60
orf48a.pep	MNIHTLLSKQ	WTLPPFLPKR	LLLSLILLX	PNAVFWLAL	LTATARPIVN	LXLYPAALLI
35	orf48-1	MNIHTLLSKQ	WTLPPFLPKR	LLLSLILLX	PNAVFWLAL	LTATARPIVN
	10	20	30	40	50	60
	70	80	90	100	110	120
40	orf48a.pep	ALPWRXVKIX	GVLAXWLAVL	FDGLMMVIQL	FPFMDLIGAI	NLVPFIXTAP
	orf48-1	ALPWRXVKIX	GVLAXWLAVL	FDGLMMVIQL	FPFMDLIGAI	NLVPFIXTAP
	70	80	90	100	110	120
45	orf48a.pep	LLYMLAMPFV	LQKAAAKTDF	RHIAACAADV	VAAGYFTGHL	SXYDRGRMAN
	orf48-1	LLYMLAMPFV	LQKAAAKTDF	RHIAACAADV	VAAGYFTGHL	SXYDRGRMAN
	130	140	150	160	170	180
50	orf48a.pep	KSQAMLYTVS	QNADFITAGL	VDPVFLPLGN	QQRAATHLNE	PKSQKILFIV
	orf48-1	KSQAMLYTVS	QNADFITAGL	VDPVFLPLGN	QQRAATHLNE	PKSQKILFIV
	190	200	210	220	230	240
55	orf48a.pep	ELQNATFAKL	LAQKXRFVSW	ESGSFPIGA	TIEGEMRELC	AYGGLRGFAL
	orf48-1	ELQNATFAKL	LAQKXRFVSW	ESGSFPIGA	TIEGEMRELC	AYGGLRGFAL
	250	260	270	280	290	300
60	orf48a.pep	CLPNRLKQEG	YATFAMHGAG	SSLYDRFSWY	PRAGFQEIKT	AENLIGKKT
	orf48-1	CLPNRLKQEG	YATFAMHGAG	SSLYDRFSWY	PRAGFQEIKT	AENLIGKKT
	310	320	330	340	350	360
65	orf48a.pep	LFGEVSAXFK	KHDKGLFYWM	TLTSHADYPE	SDIFNHLK	TEYGLPAETD
	orf48-1	LFGEVSAXFK	KHDKGLFYWM	TLTSHADYPE	SDIFNHLK	TEYGLPAETD

-280-

38

		370	380	390	400	410	420
	orf48a.pep	LFGEVSAXFKKHKDGLFYWMTLTSHADYPESDIFNHLKCKTEYGLPAETDXCRNFSLHTQ					
5	orf48-1	LFGEVSAFFKKHKDGLFYWMTLTSHADYPESDIFNHLKCKTEYGLPAETDLCRNFSLHTQ					
		370	380	390	400	410	420
	orf48a.pep	FFDQLADLIQRPEMKGTEVIIIVGDHPPVGNLNETFRYLKQGHVXWLNFKIKX					
10	orf48-1	FFDQLADLIQRPEMKGTEVIIIVGDHPPVGNLNETFRYLKQGHVXWLNFKIKX					
		430	440	450	460	470	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF48 shows 97.5% identity over a 119aa overlap with a predicted ORF (ORF48ng) from *N.*

15 *gonorrhoeae*:

	orf48.pep	MNIHTLLSKQWTLPPFLPKRLLLSLILLAPNAVEFWVLALLTATARPVNL DYLP AALLI	60
	orf48ng	MNIHALLSEQWTLPPFLPKRLLLSLILLAPNAVEFWVLALLTATARPVNL DYLP AALLI	60
20	orf48.pep	ALPWRFVKIAGVLAFWLAVLFDGLMMVQLFPFMDLIGAINLVPFILTAPAPYQIMTGL	119
	orf48ng	ALPWRFVKIAGVLAFWPAVLFDGLMMVQLFPFMDLIGAINLVPFILTAPAPYQIMTGLL	120

The ORF48ng nucleotide sequence <SEQ ID 473> was predicted to encode a protein having amino acid sequence <SEQ ID 474>:

25	1	MNIHALLSEQ WTLPPFLPKR LLSLILLILA PNAVEFWLAL LTATARPVIV
	51	LDYLP AALLI ALPWRFVKIA GVLAFWPAVL FDGLMMVQL FPFMDLIGAI
	101	NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAVKTD F RHIAVCAAVV
	151	AAARYFTGPF ELLRTGGRWQ YVQHRRLLS GSRASFRRRQ KADVLRR LGN
	201	PYASMGNGG..

30 Further work identified the complete gonococcal DNA sequence <SEQ ID 475>:

	1	ATGAATATTC	ACGCCCTGCT	CTCCGAACAA	TGGACGCTGC	CGCCATTCCT
	51	GCCGAAACGG	CTGCTGCTGT	CCCTGCTGAT	ACTGCTGGCC	CCCAATGCGG
	101	TGTTTTGGGT	TTTGGCACTG	CTGACCGCCA	CCGCCCGCCC	GATTGTCAAT
35	151	TTGGA CTACC	TTCCCGCCGC	GCTGCTGATC	GCCCTGCCTT	GGCGTTTCGT
	201	CAAAATTGCC	GGCGTATTGG	CGTTTGGCC	GGCGGTTTTG	TTTGACGGGC
	251	TGATGATGGT	GATCCAATC	TTCCCTTTTA	TGGACCTCAT	CGCGGCCATC
	301	AACCTCGTCC	CCTTCATCCT	GACCGCCCC	GCCCCTTATC	AGATAATGAC
	351	CGGGCTGTG	CTGCTGTATA	TGCTGGCGAT	GCCGTTTGTG	TTGCAAAAAG
40	401	CCGCCGTCAA	AACCGACTTC	CGACACATTG	CCGTCTGTGC	CGCCGTTGTG
	451	GCGGCAGCCG	GCTATTTTAC	CGGCCATTTG	AGTTACTACG	ACCGGGGGCG
	501	GATGGCCAAT	ATCTTCGCGC	CAACAACCTT	CTATTACGCG	aAAAGTCAGG
	551	CGATGCTCTA	CACCGTCAGC	CAGAATGCCG	ACTTTATTAC	CGCCGgcctG
	601	GTCGACCCCG	TCTTCCTCCC	CTTGGGCAAT	CAGCAGCGTG	CGGCCACGCG
	651	GCTGAGTGAG	CCGAAATCTC	AAAAAATCCT	CTTTATCGTC	GCCGAATCTT
45	701	GGGGGCTGCC	GGGCAATCCC	GAGCTTCAAA	ACGCCACTTT	TGCCAAACTG
	751	CTGGCGCAAA	AAGACCGTTT	TTCGGTTTGG	GAAAGCGGCA	GTTTTCCTT
	801	CATCGGCGCG	ACGGTCGAAG	GCGAAATGCG	CGAATTGTGC	GCCTACGGCG
	851	GTTTGC GCGG	GTTTCGCACTG	CGCCGCGCGC	CCGACGAAAA	ATTGCCCCGC
	901	TGCCTCCCCA	ACCGTTTGAA	ACAAGAAGGT	TACGCCACCT	TTGCGATGCA
50	951	CGGCGCGGGT	AGTTTCGTTT	ACGACCGCTT	CAGCTGGTAT	CCGAGGGCGG
	1001	GCTTTTCAAAA	AATCAAAACC	GCCGAAAACC	TGATCGGTAA	AAAAACCTGC
	1051	GCCATTTTCG	GCGGCGTGTG	CGACAGCGAG	CTGTTCCGGC	AAGTGTCCGC
	1101	ATTTTTCAAA	AAACACGACA	AGGGACTGTT	TTACTGGATG	ACGCTGACCA
	1151	GCCACGCCGA	CTATCCCGAA	TCCGACATTT	TCAACCACAG	GCTCAAATGC
55	1201	ACCGAATACG	GCCTGCCCGC	CGAAACCGAC	CTCTGCCGCA	ATTTCAGCCT
	1251	GCACACCCAA	TtettegACC	AACTGGCGGA	TTGATCCGA	CGCCCCGAAA
	1301	TGAAAGGCAC	GGAAGTCATC	ATCGTCGGCG	ACCATCCGCC	GCCCGTCGGC
	1351	AACCTCAATG	AAACCTTCCG	CTACCTCAAA	CAGGGACACG	TGCCTGGCT
	1401	GCACTTCAAA	ATCAAATAA			

60 This encodes a protein having amino acid sequence <SEQ ID 476; ORF48ng-1>:



-281-

1 MNIHALLSEQ WTLPPFLPKR LLSLLILLA PNAVFWVLAL LTATARPIVN  
 51 LDYLPALLI ALPWRFKIA GVLAFWPAVL FDGLMMVIQL FPFMDLIGAI  
 101 NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAVKTD F RHIIVCAAVV  
 151 AAAGYFTGHL SYDRGRMAN IFGANNFYA KSQAMLYTVS QNADFITAGL  
 5 VDPVFLPLGN QQRAATRLSE PKSQKILFIV AESWGLPGNP ELQNATFAKL  
 201 LAQKDRFSVW ESGSFFFIGA TVEGEMREL CAYGGLRGFAL RRAPDEK FAR  
 251 CLPNRLKQEG YATFAMHGAG SSLYDRFSWY PRAGFQIKT AENLIGKKT C  
 301 AIFGGVCDSE LFGEVSAFFK KHDKGLFYWM TLTSHADY P SDIFNHLK C  
 351 TEYGLPAETD LCRNFSLHTQ FFDQLADLIR RPEMKGTEVI IVGDHPPV G  
 10 401 NLNETFRYLK QGHVAWLHFK IK\*  
 451

ORG48ng-1 and ORF48-1 show 97.9% identity in 472 aa overlap:

		10	20	30	40	50	60
	orf48-1.pep	MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNL	DYLPALLI				
15	orf48ng-1	MNIHALLSEQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNL	DYLPALLI				
		10	20	30	40	50	60
	orf48-1.pep	ALPWRFKIAGVLAFWLAVLFDGLMMVIQLFPFMDLIGAINLVVPFILTAPAPYQIMTGLL					
20	orf48ng-1	ALPWRFKIAGVLAFWPAVLFDGLMMVIQLFPFMDLIGAINLVVPFILTAPAPYQIMTGLL					
		70	80	90	100	110	120
	orf48-1.pep	LLYMLAMPFVLQKAAAKTDFRHIIVCAAVVAAAGYFTGHL	SYDRGRMANIFGANNFYA				
25	orf48ng-1	LLYMLAMPFVLQKAAVKTD RHIIVCAAVVAAAGYFTGHL	SYDRGRMANIFGANNFYA				
		130	140	150	160	170	180
	orf48-1.pep	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRAATHLNEPKSQKILFIVAESWGLPANP					
30	orf48ng-1	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRAATRLSEPKSQKILFIVAESWGLPGNP					
		190	200	210	220	230	240
	orf48-1.pep	ELQNATFAKLLAQKDRFSVWESGSFPFIGATVEGEMRELCA	YGGGLRGFALRRAPDEK FAR				
35	orf48ng-1	ELQNATFAKLLAQKDRFSVWESGSFPFIGATVEGEMRELCA	YGGGLRGFALRRAPDEK FAR				
		250	260	270	280	290	300
	orf48-1.pep	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQIKTAENLIGKKTCAIFGGVCDSE					
40	orf48ng-1	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQIKTAENLIGKKTCAIFGGVCDSE					
		310	320	330	340	350	360
	orf48-1.pep	LFGEVSAFFKKHDKGLFYWM TLTSHADY P	SDIFNHLKCTEYGLPAETD	LCRNFSLHTQ			
45	orf48ng-1	LFGEVSAFFKKHDKGLFYWM TLTSHADY P	SDIFNHLKCTEYGLPAETD	LCRNFSLHTQ			
		370	380	390	400	410	420
	orf48-1.pep	FFDQLADLIR RPEMKGTEVI	IVGDHPPVGNLNETFRYLKQGHVAWLHFKIKX				
50	orf48ng-1	FFDQLADLIR RPEMKGTEVI	IVGDHPPVGNLNETFRYLKQGHVAWLHFKIKX				
		430	440	450	460	470	
	orf48-1.pep	FFDQLADLIR RPEMKGTEVI	IVGDHPPVGNLNETFRYLKQGHVAWLHFKIKX				
55	orf48ng-1	FFDQLADLIR RPEMKGTEVI	IVGDHPPVGNLNETFRYLKQGHVAWLHFKIKX				

Based on this analysis, including the presence of a putative leader sequence (double-underlined)  
 and two putative transmembrane domains (single-underlined) in the gonococcal protein, it is  
 predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be  
 useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 57**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 477>:

```

1   ..GTGAGCGGAC GTTACCGCGC TTTGGATCGC GTTTCCAAAA TCATCATCGT
51  TACTTTGAGT ATCGCCACGC TTGCCGCCGC CGGCATCGCT ATGTCGCGCG
101 GTATGCAGAT GCAGTCCGAT TTTATCGAGC CGACACCGTG GACGCTTGCC
151 GGTTTGGGCT TCCTGATCGC GCTGATGGGC TGGATGCCCG CGCCGATTGA
201 AATTTCCGCC ATCAATTCTT TGTGGGTAAC CGAAAAACAA CGCATCAATC
251 CTTCCGAATA CCGCGACGGG ATTTTGAAT TCAACGTCGG TTATATCGCC
301 AGTGCGGTTT TGGCTTTGGT TTTCTTGCA CTGGGCGC.G TAGCGCCGAA
10 351 CGGCAACGGC GA.ACAGTGC AGATGGCGGG CGGCAATAT AACGGGCAAT
401 TGATCAATAT GTACGCC..

```

This corresponds to the amino acid sequence <SEQ ID 478; ORF53>:

```

1   ..VSGRYRALDR VSKIIIVTLS IATLAAAGIA MSRGMQMOSD FIEPTPWTLA
51  GLGFLIALMG WMPAPIEISA INSLWVTEKQ RINPSEYRDG IFEFNVGYIA
15 101 SAVLALVFLA LGXVAPNGNG XTVQMAGGKY NQQLINMYA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 479>:

```

1   ATGTCCGAAC AACATATTTT GACTTGAAA AGTAAATCA ACGCATTGGG
51  TCCGGGGATC ATGATGGCTT CGGCGGCGGT CGGCGGTTCC CACCTGATTG
101 CCTCGACGCA GGCGGGCGCG CTTTACGGCT GGCAGATCGC GCTCATCATC
20 151 ATCCTGACCA ACCTCTTCAA ATACCCGTTT TTCCGCTTCA GCGCGCATTA
201 CACGCTGGAC ACGGGCAAGA GCCTGATTGA AGGTTATGCC GAGAAAAGCC
251 GCGTTTATTT GTGGGTATTC CTGATTTTGT GCATCCTCTC CGCCACGATT
301 AACGCGGGCG CGGTCGCCAT TGAACCGCC GCCATCGTCA AAATGGCGAT
351 TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCCTTG ATTATGGCAT
25 401 CCTGCCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT GGATCGCGTT
451 TCCAAAATCA TCATCGTTAC TTTGAGTATC GCCACGCTTG CCGCCGCCGG
501 CATCGCTATG TCGCGCGGTA TGCAGATGCA GTCCGATTTT ATCGAGCCGA
551 CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT GATGGGCTGG
601 ATGCCCGCGC CGATTGAAAT TTCCGCCATC AATTCTTTGT GGGTAACCGA
30 651 AAAACAACGC ATCAATCCTT CCGAATACCG CGACGGGATT TTTGATTTCA
701 ACGTCGGTTA TATCGCCAGT GCGGTTTTGG CTTTGGTTTT CTTGCACTG
751 GGCGCGTTTG TGCAATACGG CAACGGCGAA GCAGTGCAGA TGGCGGGCGG
801 CAAATATATC GGGCAATTGA TCAATATGTA CGCCGTTACC ATCGGCGGCT
851 GGTGCGGCCC GCTGGTGGCG TTTATCGCGT TTGCCTGTAT GTACGGCAGC
35 901 ACGATTACCG TCGTGGACGG CTATGCCCGT GCCATTGCCG AACCCGTGCG
951 CCTGTGCGC GGAAGAGACA AAACGGGCAA CGCCGAATTC TTTGCTTGA
1001 ATATTTGGGT GGCGGGCAGC GGTTTGGCGG TGATTTTCTG GTTTGACGGC
1051 GTAATGGCGA ATCTGCTCAA ATTTGCGATG ATTGCCGCTT TTGTGTCCGC
1101 CCCTGTGTTT GCCTGCTGTA ATTACCGTTT GGTTAAAGGT GATGAAAAAC
40 1151 ACAAACACAC ATCAGGTATG AATGCCCTTG CATTGGCAGG CTTGATTAT
1201 CTGACCGGTT TTACCGTTTT GTTCTTATTG AATTTGGCGG GAATGTTCAA
1251 ATGA

```

This corresponds to the amino acid sequence <SEQ ID 480; ORF53-1>:

```

1   MSEQHISTWK SKINALGPGI MMASAAVGGG HLIASQAGA LYGWQIALII
45 51  ILTNLFKYPF FRFSAHYTLD TGKSLIEGYA EKSRVYLWVF LILCILSATI
101 NAGAVAIVTA AIVKMAIPSL MFDAGTVAAL IMASCLIIIV SGRYRALDRV
151 SKIIIVTSLI ATLAAAGIAM SRGMQMOSDF IEPTPWTLAG LGFLIALMGW
201 MPAPIEISAI NSLWVTEKQR INPSEYRDGI FDFNVGYIAS AVLALVFLAL
251 GAFVQYNGNE AVQMAGGKYI QQLINMYAVT IGGWSRPLVA FIAFACMYGT
50 301 TITVVDGYAR AIAEPVRLLR GKDKTGNAEF FAWNIWVAGS GLAVIFWFDG
351 VMANLLKFAM IAAFVSAPVF AWLNYRLVKG DEKHKLTSGM NALALAGLIY
401 LTGFTVLELL NLAGMFK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

55 ORF53 shows 93.5% identity over a 139aa overlap with an ORF (ORF53a) from strain A of *N. meningitidis*:

-283-

					10	20	30
	orf53.pep				VSGRYRALDRVSKIIIVTL	SIATLAAAGIA	
5	orf53a	AAIVKMAIPSLMFDAGTVAALIMASCLIIILV	S	G	R	Y	R
		110	120	130	140	150	160
		40	50	60	70	80	90
	orf53.pep	MSRGMQMOSDFIEPTPWTLAGLGLFI	A	L	M	G	W
10	orf53a	MSRGMQMOSDFIEPTPWTLAGLGLFI	A	L	M	G	W
		170	180	190	200	210	220
		100	110	120	130	139	
	orf53.pep	IFEFNVG	Y	I	A	S	A
15	orf53a	IFDFNV	G	Y	I	A	S
		230	240	250	260	270	280
20	orf53a	AFIAFACMYGTTITVVDGYARAIAEPVRL	L	R	G	K	D
		290	300	310	320	330	340

The complete length ORF53a nucleotide sequence <SEQ ID 481> is:

	1	ATGTCCGAAC	AACATATTTTC	GA	CTTGGAAA	AGTAAATCA	ACGCATTGGG
	51	ACCGGGGATT	ATGATGGCTT	CG	GCGGCGGT	CGGCGGTT	CGCACCTGATTG
25	101	CCTCGACGCA	GGCGGGCGCG	CT	TACGGCT	GGCAGATCGC	GCTCATCATC
	151	ATCCTGACCA	ACCTCTTCAA	AT	ACCGTTT	TTCCGCTTCA	GCGCGCATT
	201	CACGCTGGAC	ACGGGCAAGA	GC	CTGATTGA	AGGTTATGCC	GAGAAAAGCC
	251	GCGTTTATTT	GTGGGTATTC	CT	GATTTTGT	GCATCCTCTC	CGCCACGATT
	301	AACGCGGGCG	CGGTCGCCAT	TG	TAACCGCC	GCCATCGTCA	AAATGGCGAT
	351	TCCCTCGCTG	ATGTTTGATG	CC	GGCACGGT	TGCCGCTTG	ATTATGGCAT
30	401	CCTGCCTGAT	TATTTTGGTG	AG	CGGACGTT	ACCGCGCTTT	GGATCGCGTT
	451	TCCAAAATCA	TCATCGTTAC	TT	TGAGTATC	GCCACGCTTG	CCGCCGCCGG
	501	CATCGCTATG	TCGCGCGGTA	TG	CAGATGCA	GTCCGATTTT	ATCGAGCCGA
	551	CACCGTGGAC	GCTTGCCGGT	TT	GGGCTTCC	TGATCGCGCT	GATGGGCTGG
	601	ATGCCCGCGC	CGATTGAAAT	TT	CCGCCATC	AATTCCTTGT	GGGTAAACCGA
35	651	AAAACAACGC	ATCAATCCTT	CC	GAATACCG	CGACGGGATT	TTTGATTTC
	701	ACGTCGGTTA	TATCGCCAGT	GC	GGTTTGG	CTTTGGTTT	CCTTGCACTG
	751	GGCGCGTTTG	TGCAATACGG	CA	ACGGCGAA	GCAGTGCAGA	TGGCGGGCGG
	801	CAAATATATC	GGGCAATTGA	TC	AATATGTA	CGCCGTTACC	ATCGGCGGCT
	851	GGTCGCGCCC	GCTGGTGGCG	TT	TATCGCGT	TTGCCTGTAT	GTACGCAACG
40	901	ACGATTACCG	TTGTGGACGG	CT	ATGCCCGT	GCCATTGCCG	AACCCGTGCG
	951	CCTGCTGCGC	GGAAAAGACA	AA	ACGGGCAA	CGCCGAATTC	TTTGCCTGGA
	1001	ATATTTGGGT	GGCGGGCAGC	GG	TTTGGCGG	TGATTTTCTG	GTTTGACGGC
	1051	GTAATGGCGA	ATCTGCTCAA	ATT	TGCGATG	ATTGCCGCTT	TTGTGTCCGC
	1101	CCTGTGTGTT	GCCTGCTGA	ATT	ACCGTTT	GGTCAAAGGT	GATGAAAAAC
45	1151	ACAAACTCAC	ATCAGGTATG	AAT	GCCCTTG	CATTGGCAGG	CTTGATTAT
	1201	CTGACCGGTT	TTACCGTTTT	GTT	CTTATTG	AATTTGCGCG	GAATGTTCAA
	1251	ATGA					

This encodes a protein having amino acid sequence <SEQ ID 482>:

	1	MSEQHISTWK	SKINALGPGI	MMASAAVGG	SHLI	ASTQAGA	LYGWQIALII
50	51	ILTNLFKYPF	FRFSAHYTLD	TG	KS	LIEGYA	EKSRVYLWVF
	101	NAGAVAIVTA	AI	V	K	MAIPSL	MFDAGTVAAL
	151	SKIIIVTL	SI	ATL	AAAGIAM	SRGMQMOSDF	IEPTPWTLAG
	201	MPAPIEISAI	NSLWVTEKQ	R	INPSEYR	DGI	FD
	251	FAFVQYNGE	AVQ	MAGGKYI	GQLINMY	AVT	IGGWSRPLVA
55	301	TITVVDGYAR	AI	AE	PVRL	LR	GKDKTGNAEF
	351	VMANLLKFAM	IA	AFV	SAPVF	AWLN	YRLVK
	401	LTGFTVLFLL	NLAGMFK*				

ORF 53a shows 100.0% identity in 417 aa overlap with ORF53-1:

		10	20	30	40	50	60
60	orf53a.pep	MSEQHISTWKS	K	I	N	A	L
	orf53-1	MSEQHISTWKS	K	I	N	A	L
		10	20	30	40	50	60
65		70	80	90	100	110	120

-284-

	orf53a.pep	FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTAIVKMAIPSL
	orf53-1	FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTAIVKMAIPSL
5		70 80 90 100 110 120
	orf53a.pep	130 140 150 160 170 180
	orf53-1	MFDAGTVAALIMASCLIIILVSGRYRALDRVSKIIIVTLSIATLAAAGIAMSRGMQMQSDF
10		130 140 150 160 170 180
	orf53a.pep	190 200 210 220 230 240
	orf53-1	IEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDGIFDFNVGYIAS
15		190 200 210 220 230 240
	orf53a.pep	250 260 270 280 290 300
	orf53-1	AVLALVFLALGAFVQYGNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVAFIAFACMYGT
20		250 260 270 280 290 300
	orf53a.pep	310 320 330 340 350 360
	orf53-1	TITVVDGYARAIAEPVRLLRGKDKTGNAEFFAWNIWVAGSGLAVIFWFDGVMANLLKFAM
25		310 320 330 340 350 360
	orf53a.pep	370 380 390 400 410
	orf53-1	IAAFVSAPVFAWLNYRLVKGDEKHKLTSGMNALALAGLIYLTGFTVFLNLAGMFKX
30		370 380 390 400 410
	orf53a.pep	IAAFVSAPVFAWLNYRLVKGDEKHKLTSGMNALALAGLIYLTGFTVFLNLAGMFKX
35		370 380 390 400 410

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF53 shows 92.1% identity over a 139aa overlap with a predicted ORF (ORF53ng) from *N. gonorrhoeae*:

40	orf53.pep	VSGRYRALDRVSKIIIVTLSIATLAAAGIA	30
	orf53ng	AAIVKMAIPSLMFDAGTVAALIMASCLIIILVSGRYRALDRVSKIIIVTLSIATLAAAGIA	91
	orf53.pep	MSRGMQMQSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	90
45	orf53ng	MSRGMQMQPDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	151
	orf53.pep	IFEFNVGYIASAVLALVFLALGXVAPNGNGXTVQMAGGKYNGQLINMYA	139
50	orf53ng	IFDFNVGYIASAVLALVFLALGAFVQYGNGEAVQMGGGKYIGQLINMYAVTIGGGSRPLV	211

An ORF53ng nucleotide sequence <SEQ ID 483> was predicted to encode a protein having amino acid sequence <SEQ ID 484>:

55	1	MPKKSCVYLW VFLILCIASA TINAGAVAIV TAAIVKMAIP SLMFDAGTVA
	51	ALIMASCLII LVSGRYRALD RVSKIIIVTL SIATLAAAGI AMSRGMQMQP
	101	DFIEPTPWTLAGLGFLIALMGWMPAPIEIS AINSLWVTEK QRINPSEYRD
	151	GIFDFNVGYI ASAVLALVFL ALGAFVQYGN GEAVQMGGGK YIGQLINMYA
	201	VTIGGGSRPL VAFIAFACMY GAASTVVDGY ARAIAEPVRL LRGKDKTARP
	251	IVLLEKLGR HRFRGDFLV*

Further analysis revealed further partial DNA gonococcal sequence <SEQ ID 485>:

60	1	..aagaAAAGCT GCGTTTATTT GTGGGTTTTT TTGATTTTGT GTATCGCCTC
	51	CGCCACGATT AACCGGGGCG CGGTGCGCAT TGTAACCGCC GCCATCGTCA
	101	AAATGGCGAT TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCGCTTG

-285-

151 ATTATGGCAT CCTGCCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT  
 201 GGATCGTGTT TCCAAATCA TCATTGTTAC TTTGAGCATC GCCACGCTTG  
 251 CCGCCGCCGG CATCGCTATG TCGCGCGGTA TGCAGATGCA GCGCGATTTT  
 301 ATCGAGCCGA CACCGTGGAC GCTTGGCCGGT TTGGGCTTCC TGATCGCGCT  
 351 GATGGGCTGG ATGCCCGCGC CGATCGAAAT TTCCGCCATC AATTCTTTGT  
 401 GGGTAACCGA AAAACAACGC ATCAATCCTT CTGAATACCG CGACGGGATT  
 451 TTCGATTTC ACGTCGGTTA TATCGCagT GCGGTTTTGG CTTTGGTTTT  
 501 CCTTGCACTG GCGCGTGTG TGCAATACGG CAACGGCGAA GCAGTGCAGA  
 551 TGGCGGGCGG CAAATATATC GGGCAATTGA TTAATATGTA TGCCGTAACC  
 601 ATCGGCGGCT GGTCTCGTCC GCTGGTGGCG TTTATCGCGT TTGCCTGTAT  
 651 GTACGGCACG ACGATTACCG TTGTGGACGG TTATGCGCGT GCCATTGCCG  
 701 AACCCGTGCG CCTGCTGCGC GGCAGGGATA AAACGGCAA CGCCGAGTTG  
 751 TTtgcctGGA ATATTTGGGT GCGGGGCGAG GGTTTGGCGG TGATTTTCTG  
 801 GTTTGACggc gcaatGGCgG AActgcTCAA ATTTGCGATG ATtgcgcCT  
 851 TTGTGTCCG CCCTGTGTTC GCCTGGCTCA ACTACCGCCT CGTCAAAGGG  
 901 GACAAACGCC ACAGGCTTAC CGCCGGTATG AACGCCCTTG CCATTGTCGG  
 951 CCTGCTCTAC CTGGCCGGGT TTGCCGTTTT GTTCCTGTTG AACCTTACCG  
 1001 GACTTTTGGC ATAG

This corresponds to the amino acid sequence <SEQ ID 486; ORF53ng-1>:

1 ..KKSCVYLWVF LILCIASATI NAGAVAIVTA AIVKMAIPSL MFDAGTVAAL  
 51 IMASCLIIIV SGRYRALDRV SKIIIVTLSI ATLAAAGIAM SRGMQMOPDF  
 101 IEPTPWTLAG LGFLIALMGW MPAPIEISAI NSLWVTEKQR INPSEYRDGI  
 151 FDFNVGYIAS AVLALVFLAL GAFVQYNGE AVQMAGGKYI GQLINMYAVT  
 201 IGGWSRPLVA FIAFACMYGT TITVVDGYAR AIAEPVRLLR GRDKTGNAEL  
 251 FAWNIWVAGS GLAVIFWFDG AMAELLKFAM IAAFVSAPVF AWLNYRLVKG  
 301 DKRHRLTAGM NALAIVGLLY LAGFAVLFL NLTGLLA\*

ORF53ng-1 and ORF53-1 show 94.0% identity in 336 aa overlap:

		60	70	80	90	100	110
30	orf53-1.pep	ILTNLFKYPFFRFS	AHYTLDTGKS	LIEGYAEKSRVYLWV	FLILCILSATINAGAVAIVTA		
	orf53ng-1						
					10	20	30
		120	130	140	150	160	170
35	orf53-1.pep	AIVKMAIPSLMFDAGTVAAL	IMASCLIIIVSGRYRALDRV	SKIIIVTLSIATLAAAGIAM			
	orf53ng-1						
		40	50	60	70	80	90
40	orf53-1.pep	SRGMQMOSDFIEPTPWTL	LAGLGFLIALMGWMPAPIE	ISAINSLWVTEKQRINPSEYRDGI			
	orf53ng-1						
		100	110	120	130	140	150
45	orf53-1.pep	FDFNVGYIASAVLALVFL	ALGAFVQYNGEAVQMAGGKYI	GQLINMYAVTIGGWSRPLVA			
	orf53ng-1						
50		160	170	180	190	200	210
	orf53-1.pep	FIAFACMYGTTITVVDGYAR	AIAEPVRLLRGDKDTGNAE	FFAWNIWVAGSGLAVIFWFDG			
	orf53ng-1						
55		220	230	240	250	260	270
	orf53-1.pep	VMANLLKFAMIAAFVSAPVF	AWLNYRLVKGDEKHKLTSGM	NALALAGLIYLTGFTVLFL			
60	orf53ng-1						
		280	290	300	310	320	330
65	orf53-1.pep	NLAGMFKX					
	orf53ng-1	:	NLTGLLAX				

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 58

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 487>:

```

1  ..TTGCGGGAAA CGGCATATGT TTTGGATAGT TTTGATCGTT ATTTTGTGTG
51  TGCGCTTGCC GGCTTGTTTT TTGTCCGCGC ACAATCCGAA CGCGAGTGGA
101 TGCGCGAGGT TTCTGCGTGG CAGGAAAAGA AAGGGGAAAA ACAGGCGGAG
151 CTGCCTGAAA TCAAAGACGG TATGCCCGAT TTTCCCGAAC TTGCCCTGAT
201 GCTTTTCCAC GCCGTCAAAA CGGCAGTGTA TTGGCTGTTT GTCGGTGTCTG
251 TCCGTTTCTG CCGAAACTAT CTGGCGCAGC AATCCGAACC GGACAGGCC
301 GTTCCGCCT..

```

This corresponds to the amino acid sequence <SEQ ID 488; ORF58>:

```

1  ..LRETAYVLDS FDRYFVVALA GLFFVRAQSE REWMREVSAA QEKKGEKQAE
51  LPEIKDGMPP FPELALMLFH AVKTAVYWLF VGVVRFRCRNY LAHESEPDRL
101 VPP..

```

Further work revealed the complete nucleotide sequence <SEQ ID 489>:

```

20  1  ATGTTTTGGA TAGTTTTGAT CGTTATTTTG TTGCTTGC GC TTGCCGGCTT
51  GTTTTTTGTC CGCGCACAAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG
101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAATCAAAA
151 GACGGTATGC CCGATTTTCC CGAACTTGCC CTGATGCTTT TCCATGCCGT
201 CAAAACGGCA GTGTATTGGC TGTTTGTCGG TGTCGTCCGT TTCTGCCGAA
25  251 ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT
301 GCAAACCGTG CGGATGTTCC GACCGCATCC GACGGATATT CAGACAGTGG
351 AAACGGGACG GAAGAAGCGG AAACGGAAGA AGCAGAAGCT GCGGAGGAAG
401 AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC
451 ATCCCATTCTG ACCGGAGTAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA
30  501 AATTTGCCCC GTCCGTCCGG TTTTAAAGA AATCACTTTG GAAGAAGCAA
551 CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC
601 GATGCATTTG AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCGGA
651 TACCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC
701 AACGCACGTA TTCCCATATG TTCGATGCGG ACAAAGAAGC GTTTTCCGAG
35  751 TCTGCGGATT ACGGATTTGA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC
801 CTTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCCACCGTC
851 ATGCAGGGCA GGGGAAAGGG CAGGCGGAGG CAAAATCCCC GGATGTTTCC
901 CAAGGGCAGT CCGTTTCAGA CGGCACGGCC GTCCGCGATG CCCGCCGCCG
951 CGTTTCCGTC AATTTGAAAG AACCGAACAA GGCAACGGTT TCTGCGGAGG
40  1001 CGCGAATTTT TCGCCTGATT CCGGAAAGTC AGACGGTTGT CGGGAAACGG
1051 GATGTCGAAA TGCCGTCTGA AACCAGAAAT GTTTTCACGG AAACCGTTTC
1101 GTCTGTGGGA TACGGCGGTC CGGTTTATGA TGAAACTGCC GATATCCATA
1151 TTGAAGAACC TGCCGCGCCC GATGCTTGGG TGGTCGAACC ACCCGAAGTG
45  1201 CCGAAAGTTC CCATGACCGC AATCGATATT CAGCCGCCGC CTCCCGTATC
1251 GGAAATCTAC AACCGTACCT ATGAACCGCC GTCAGGATTC GAGCAGGTGC
1301 AACGCAGCCG CATTGCCGAG ACCGACCATC TTGCCGATGA TGTTTGAAT
1351 GGAGGTTGGC AGGAGGAAAC CGCCGCTATT GCGGATGACG GCAGTGAAGG
1401 TGCGGCAGAG CGGTCAAGCG GGCAATATCT GTCGGAACCC GAAGCGTTTCG
50  1451 GGCATGACAG TCAGGCGGTT TGTCCGTTTG AAAATGTGCC GTCTGAACGC
1501 CCGTCCTGCC GGTATCGGA TACGGAAGCG GATGAAGGGG CGTTCCCATC
1551 TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
1601 TGCTTCCGCT GTTCAATCCC GAGGCGACGC AAACCGAAGA AGAACTGTTG
1651 GAAAAACAGCA TCACCATCGA AGAAAAATTG GCGGAGTTCA AAGTCAAGGT
55  1701 CAAGGTTGTC GATTCTTATT CCGGCCCGGT AATTACGCGT TATGAAATCG
1751 AACCAGATGT CGGCGTGC GC GAATTTCCG TTCTGAATCT GGAAAAAGAT
1801 TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTTGTCT AAACCATCCC
1851 CGGCAAAACC TGCATGGGTT TGGAACTTCC GAACCCGAAA CGCCAAATGA

```

-287-

1901 TACGCCTGAG CGAAATCTTC AATTCGCCCC AGTTTGCCGA ATCCAAATCC  
 1951 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCGTAAC  
 2001 CGACTTGGGA AAAGCACCAG ATTTGTTGGT TGCCGGCACG ACCGGTTCGG  
 5 2051 GCAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTTCAAAGCC  
 2101 GCGCCGGAAG ACGTGCCTAT GATTATGATC GATCCGAAAA TGCTGGAATT  
 2151 GAGCATTTAC GAAGGCATCC CGCACCTGCT CGCCCCCTGTC GTTACCGATA  
 2201 TGAAGCTGGC GGCAAACGCG CTGAACTGGT GTGTTAACGA AATGGAAAAA  
 2251 CGCTACCGCC TGATGAGCTT TATGGGCGTG CGTAATCTTG CGGGCTTCAA  
 10 2301 TCAAAAAATC GCCGAAGCCG CAGCAAGGGG AGAAAAATC GGCAATCCGT  
 2351 TCAGCCTCAC GCCCCACGAT CCCGAACCTT TGGAAAAACT GCCGTTTATC  
 2401 GTGGTCGTGG TCGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA  
 2451 AATCGAAGAA CTGATTGCCC GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA  
 2501 TCCATTTGAT TCTTGCCACA CAACGCCCCA GCGTCGATGT CATCACGGGT  
 15 2551 CTGATTAAGG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCCAGCAA  
 2601 AATCGACAGC CGCACGATTC TCGACCAAAT GGGCGCGGAA AACCTGCTCG  
 2651 GTCAGGGCGA TATGCTGTTT CTGCTGCCGG GTACTGCCTA TCCGCAGCGC  
 2701 GTTCACGGCG CGTTTGCCCT GGATGAAGAG GTGCACCGCG TGGTCGAATA  
 2751 TTTGAAACAG TTTGGCGAAC CGGACTATGT TGACGATATT TTGAGCGCGC  
 20 2801 GCGGCAGCGA AGAGCTGCCC GGCATCGGGC GCAGCGGCGA CGACGAAACC  
 2851 GATCCGATGT ACGACGAGGC CGTATCCGTT GTCCTGAAAA CGCGCAAGC  
 2901 CAGCATTTTC GCGGTACAGC GCGCCTTGGC TATCGGCTAC AACCGCGCGC  
 2951 CGCGTCTGAT TGACCAGATG GAGGCGGAAG GCATTGTGTC CGCACCAGAA  
 3001 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTGGACAATG CTTGA

This corresponds to the amino acid sequence <SEQ ID 490; ORF58-1>:

25 1 MFWIVLIVIL LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQAE LPEIK  
 51 DGM PDPPELA LMLFHAVKTA VYWLFGVVFR FCRNYLAHES EPDRPVPPAS  
 101 ANRADVPTAS DGYSDSGNGT EEAETEEAEA AEEEEADTED IATAVIDNRR  
 151 IPFDRSIAEG LMPSESEISP VRPVFKEITL EEATRALNSA ALRETKKRYI  
 201 DAFEKNETAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSHM FDADKEAFSE  
 30 251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FHRHAGQGKG QAEAKSPDVS  
 301 QGQSVSDGTA VRDARRRVSV NLKEPNKATV SAEARISRLI PESQTVVGKR  
 351 DVEMPSETEN VETETVSSVG YGGPVYDETA DIHIEEPAAP DAWVVEPPEV  
 401 PKVPMTAIDI QPPPPVSEIY NRTYEPSPGF EQVQRSRIAE TDHLADDVLN  
 451 GGWQEETA AI ADDGSEGAAE RSSGQYLSET EAFGHDSQAV CPFENVPSER  
 35 501 PSCRVSDETA DEGAFPSSEET GAVSEHLPTT DLLLPLFPNP EATQTEEELL  
 551 ENSITIEEKL AEFKVKVKV DSYS GPVITR YEIEPDVGV GNSVLNLEKD  
 601 LARSLGVASI RVVETIPGKT CMGLELPNPK RQMIRLSEIF NSPEFAESKS  
 651 KLTLALGQDI TGQPVVTDLG KAPHLVAGT TSGSKSVGVN AMILSMLFKA  
 701 APEDVRMIMI DPKMLELSIY EGIPHLAPV VTDMLAANA LNWCVNEMEK  
 40 751 RYRLMSFMGV RNLAGFNQKI AEAARGEKI GNPFSLTPDD PEPEKL PFI  
 801 VVVVDEFADL MMTAGKKIEE LIARLAQKAR AAGIHLILAT QRPSVDVITG  
 851 LIKANIPTRI AFQVSSKIDS RTILDQMGAE NLLGQDMLF LLPGTAYPQR  
 901 VHGAFADEE VHRVVEYLKQ FGEPDYDDI LSGGGSEELP GIGRSGDDET  
 951 DPMYDEAVSV VLKTRKASIS GVQRALRIGY NRAARLIDQM EAEGIVSAPE  
 45 1001 HNGNRTILVP LDNA\*

Computer analysis of this amino acid sequence predicts the indicated transmembrane region, and also gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF58 shows 96.6% identity over a 89aa overlap with an ORF (ORF58a) from strain A of *N.*

50 *meningitidis*:

orf58.pep      10            20            30            40            50            60  
 LRETAYVLDSFD RYFVVALAGLFFVRAQSEREWMR EVSAWQEKKG EKQAE LPEIKDGM P  
 orf58a                    ::  
 55                    MFWIVLIVILLALAGLFFVRAQSEREWMR EVSAWQEKKG EKQAE LPEIKDGM P  
                          10            20            30            40            50  
 orf58.pep                    70            80            90            100  
 FP ELALMLFHAVKTAVYWLFGVVFR FCRNYLAHES EPDRPVPP  
 60 orf58a                    FP ELALMLFHAVKTAVYWLFGVVFR FCRNYLAHES EPDRPVPPASANRADVPTASDGYSD  
                          60            70            80            90            100            110

The complete length ORF58a nucleotide sequence <SEQ ID 491> is:

```

1  ATGTTTTGGA TAGTTTTGAT CGTTATTTTG TTGCTTGCGC TTGCCGGCTT
51  GTTTTTTGTC CGCGCACAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG
101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA
5   GACGGTATGC CCGATTTTCC CGAACTTGCC CTGATGCTTT TCCATGCCGT
201 CAAAACGGCA GTGTATTGGC TGTTTGTCGG TGTCGTCCGT TTCTGCCGAA
251 ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT
301 GCAAATCGTG CGGATGTTCC GACCGCATCC GACGGATATT CAGACAGTGG
351 AAACGGGACG GAAGAAGCGG AAACGGAAGA AGCAGAAGCT GCGGAGGAAG
10  AGGTGCCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC
451 ATCCCATTCG ACCGGAGTAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA
501 AATTTCGCCC GTCCGTCCGG TTTTAAAGGA AATCACTTGG GAAGAAGCAA
551 CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC
601 GATGCATTTG AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCCGA
15  TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC
701 AACGCACGTA TTCCCGTATG TTCGATGCGG ACAAGAAGC GTTTTCCGAG
751 TCTCGGATT ACGGATTTGA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC
801 CTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCCGCCGTC
851 ATGAGGGGCA GGNAAAGGG CAGGCGGAGG CNAATCCCC GGATGTTTCC
901 CAAGGGCAGT CCGTTTCAGA CGGCACAGCC GTCCGCGATG CCNGCCGCCG
951 CGTTTCCGTC AATTTGAAAG AACCGAACAA GGCAACGGTT TCTCGGAGG
1001 CGCGGATTTC CCGCTGATT CCGGAAAGTC GGACGGTTGT CGGGAACCGG
1051 GATGTCGAAA TGCCGTCTGA AACCGAAAAT GTTTTCACGG AAANTGTTTC
1101 GTCTGTGGGA TACGGCGNTC CGGTTTATGA TGAAACTGCC GATATCCATA
25  TTGAAGAACC TGCCGCGCCC GATGCTTGGG TGGTCGAACC ACCCGAAGTG
1201 CCGAAAGTTC CCATGCCCGC AATNGATATT CCGCCCGCGC CTCCCGTATC
1251 GGAATCTAC AACCGTACCT ATGAACCGCC GGCAGGATTC GAGCAGTGC
1301 AACGCAGCCG CATTGCCGAA ACCGATCATC TTGCCGATGA TGTTTTGAAT
1351 GGAGGTTGGC AGGAGGAAAC CGCCGCTATT GCGAATGACG GCAGTGAGGG
30  TGTGGCAGAG CGGTCAAGCG GGCAATATTT GTCGGAACC GAAGCGTTTCG
1451 GGCATGACAG TCAGGCGGTT TGTCGGTTTG AAAATGTGCC GTCTGAACGC
1501 CGCTCCCGCC GGCATNNGA TACGGAAGCG GATGAAGGG CGTTCGAATC
1551 TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
1601 TGCCGCGGCT GTTCAATCCC GGGGCGACGC AAACCGAAGA AGANCTGTTG
35  GANAACAGCA TCACCATCGA AGAAAAATNG GCGGAGTTCA AAGTCAAGGT
1701 CAAGGTGTGC GATTCTTATT CCGGCCCCGT GATTACGCGT TATGAAATCG
1751 AACCCGATGT CGCGGTGCGC GGCAATTCCG TTCTAAATCT GGAAAAAGAN
1801 TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTTGTCG AAACCATCCT
1851 CGGCAAAACC TGTATGGGTT TGGAACTTCC GAACCCGAAA CGCCAATGA
40  TACGCTGAG CGAAATCTTC AATTCGCCCG AGTTTGCCGA ATCCAATCC
1951 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCGTAAC
2001 CGACTTGGGC AAAGCACCGC ATTTGTTGGT TGCCGGCAGC ACCGGTTCCG
2051 GCAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTTCAAAGCC
2101 GCGCCGGAAG ACGTGCATAT GATTATGATC GATCCGAAAA TGCTGGAATT
45  GAGCATTTAC GAAGGCATCC CGCACCTGCT CGCCCTGTC GTTACCGATA
2201 TGAAGCTGGC GGCAAACGCG CTGAAGTGGT GTGTTAACGA AATGGAAAAA
2251 CGTACCGCC TGATGAGCTT TATGGGCGTG CGCAATCTTG CCGGTNTCAA
2301 TCAAAAAATC GCCGAAGCCG CAGCAAGGGG GGAGAAAATC GGCAACCCGT
2351 TCAGCTCAC GCCGACAAT CCCGAACCTT TGGANAAATT GCCGTTTATC
50  GTGGTCGTGG TTGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA
2451 AATCGAAGAA CTGATTGCCG GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA
2501 TCCATCTTAT CCTTGCCACA CAACGCCCA GTGTCGATGT CATCACGGGT
2551 CTGATTAAGG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCCAGCAA
2601 AATCGACAGC CGCAGGATTC TTGACCAAAT GGGTGCGGAA AACCTGCTCG
55  GGCAGGGCGA TATGCTGTTT CTGCCGCCG GTACGGCCTA TCCGACGCG
2701 GTTACGGCG CGTTGCCTC GGATGAAGAG GTGCACGCG TGGTCGAATA
2751 TCTGAAACAG TTTGGCGAAC CGGACTATGT TGACGATATN TTGAGCGCGC
2801 GTATGTCCGA CGATTGCTG GGAATCAGCC GGAGCGGCGA CCGCGAAACC
2851 GATCCGATGT ACCAGGAGGC CGTGTCTGTT GTTTTGAAAA CGCGCAAAGC
60  CAGCATTTCT GCGGTGCAGC GCGCATTCG TATCGGCTAT AATCGCGCCG
2951 CCGTCTGAT TGACCAGATG GAGGCGGAAG GCATTGTGTC CGCACCGGAA
3001 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTNGACAATG CTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 492>:

```

65  1  MFIVLVIVL LLAGLFFV RAQSEREWMR EVSAWQEKKG EKQALPEIK
51  DGMPDFPELA LMLFHAVKTA VYWLFGVVR FCRNYLAHES EPDRPVPAS
101 ANRADVPTAS DGYSDSNGT EEAETEEAEA AEEEEADTED IATAVIDNRR
151 IPFDRSIAEG LMPSESEISP VRPVFKEITL EATRNLNSA ALRETKKRYI
201 DAFEKNETAV PKVRVSDTPM EGLQIIIGLDD PVLQRTYSRM FDADKEAFSE

```



-289-

	251	SADYGFEPYF	EKQHPSAFSA	VKAENARNAP	FRRHAGQGKG	QAEAKSPDVS
	301	QGQSVSDGTA	VRDAXRRVSV	NLKEPNKATV	SAEARISRLI	PESRTVVGKR
	351	DVEMPSETEN	VFTEXVSSVG	YGXPVYDETA	DIHIEEPAP	WDAWVVEPEV
5	401	PKVMPAXDI	PPPPPVSEIY	NRTYEPAGF	EQVQSRIAE	TDHLADDVLN
	451	GGWQEETA	ANDGSEGVAE	RSSGQYLSET	EAFGHDSQAV	CPFENVPSER
	501	PSRRAXDTEA	DEGAFQSEET	GAVSEHLPTT	DLLLPPLFNP	GATQTEEXLL
	551	XNSITIEEKX	AEFKVKVKV	DSYSGPVITR	YEIEPDVGVR	GNSVLNLEKX
	601	LARSLGVASI	RVVETILGKT	CMGLELPNPK	RQMIRLSEIF	NSPEFAESKS
10	651	KLTLALGQDI	TGQPVVTDLG	KAPHLVAGT	TGSGKSVGVN	AMILSMLFKA
	701	APEDVRMIMI	DPKMLELSIY	EGIPHLLAPV	VTDMKLAANA	LNWCVNEMEK
	751	RYRLMSFMGV	RNLAGXNQKI	AEAAARGEKI	GNPFSLTDPN	PEPLXKLPFI
	801	VVVVDEFADL	MMTAGKKIEE	LIARLAQKAR	AAGIHLILAT	QRPSVDVITG
	851	LIKANIPTRI	AFQVSSKIDS	RTILDQMGAE	NLLGQGDMLF	LPPGTAYPQR
15	901	VHGAFASDEE	VHRVVEYLKQ	FGEPTYVDDX	LSGGMSDDLL	GISRSGDGET
	951	DPMYDEAVSV	VLKTRKASIS	GVQRALRIGY	NRAARLIDQM	EAEGIVSAPE
	1001	HNGNRTILVP	XDNA*			

ORF58a and ORF58-1 show 96.6% identity in 1014 aa overlap:

		10	20	30	40	50	60
20	orf58a.pep	MFWIVLIVILLALAGLFFVRAQSEREWREVS	SAWQEKKGKQAE	LPEIKDGM	PDFPELA		
	orf58-1	MFWIVLIVILLALAGLFFVRAQSEREWREVS	SAWQEKKGKQAE	LPEIKDGM	PDFPELA		
		10	20	30	40	50	60
25	orf58a.pep	LMLFHAVKTAVYWL	FGVVRFCRNYLA	HESEPDRPVPPAS	ANRADVPTASD	GYSDSGNGT	
	orf58-1	LMLFHAVKTAVYWL	FGVVRFCRNYLA	HESEPDRPVPPAS	ANRADVPTASD	GYSDSGNGT	
		70	80	90	100	110	120
30	orf58a.pep	EEAETEEAEAAEEEA	ADTEDIATAVIDN	RRIPFDRSIAEGL	MPSESEISPV	RPVFKEITL	
	orf58-1	EEAETEEAEAAEEEA	ADTEDIATAVIDN	RRIPFDRSIAEGL	MPSESEISPV	RPVFKEITL	
		130	140	150	160	170	180
35	orf58a.pep	EEATRALNSAALRE	TKKRYIDAFEKNE	TAVPKVRVSDTP	MEGLQIIIGLDD	PVLQRTYSRM	
	orf58-1	EEATRALNSAALRE	TKKRYIDAFEKNE	TAVPKVRVSDTP	MEGLQIIIGLDD	PVLQRTYSHM	
		190	200	210	220	230	240
40	orf58a.pep	FDADKEAFSESADY	GFEPYFEKQHPSA	FSAVKAENARNAP	FRRHAGQGKGQ	AEAKSPDVS	
	orf58-1	FDADKEAFSESADY	GFEPYFEKQHPSA	FSAVKAENARNAP	FRRHAGQGKGQ	AEAKSPDVS	
		250	260	270	280	290	300
45	orf58a.pep	QGQSVSDGTAVRD	AXRRVSVNLKEPN	KATVSAEARISRLI	PESRTVVGKR	DVEMPSETEN	
	orf58-1	QGQSVSDGTAVRD	AXRRVSVNLKEPN	KATVSAEARISRLI	PESRTVVGKR	DVEMPSETEN	
		310	320	330	340	350	360
50	orf58a.pep	VFTEXVSSVGYGXP	VYDETADIHIEEP	APADAWVVEPPEV	PKVMPAXDI	PPPPPVSEIY	
	orf58-1	VFTETVSSVGYGXP	VYDETADIHIEEP	APADAWVVEPPEV	PKVMPMTAID	IQPPPPVSEIY	
		370	380	390	400	410	420
55	orf58a.pep	NRTYEPAGFEQVQR	SRIAETDHLADDV	LNGGWQEETA	AIANDGSEGVA	ERSSGQYLSET	
	orf58-1	NRTYEPAGFEQVQR	SRIAETDHLADDV	LNGGWQEETA	AIADDGSEGVA	ERSSGQYLSET	
		430	440	450	460	470	480
60	orf58a.pep	EAFGHDSQAVCPF	ENVPSERPSRRAX	DTEADGAFQSEET	GAVSEHLPTD	LLLPPLFNP	
	orf58-1	EAFGHDSQAVCPF	ENVPSERPSRRAX	DTEADGAFQSEET	GAVSEHLPTD	LLLPPLFNP	
		490	500	510	520	530	540
65	orf58a.pep	EAFGHDSQAVCPF	ENVPSERPSRRAX	DTEADGAFQSEET	GAVSEHLPTD	LLLPPLFNP	
	orf58-1	EAFGHDSQAVCPF	ENVPSERPSRRAX	DTEADGAFQSEET	GAVSEHLPTD	LLLPPLFNP	
		490	500	510	520	530	540
70	orf58a.pep	EAFGHDSQAVCPF	ENVPSERPSRRAX	DTEADGAFQSEET	GAVSEHLPTD	LLLPPLFNP	
	orf58-1	EAFGHDSQAVCPF	ENVPSERPSRRAX	DTEADGAFQSEET	GAVSEHLPTD	LLLPPLFNP	
		490	500	510	520	530	540

-290-

		550	560	570	580	590	600
	orf58a.pep	GATQTEEXLLXNSITIEEKXAEFKVKVVD	SYSGPVITRYEIEPDV	GVGRNSVLNLEKX			
5	orf58-1	EATQTEEELENSITIEEKLAEFKVKVVD	SYSGPVITRYEIEPDV	GVGRNSVLNLEKD			
		550	560	570	580	590	600
	orf58a.pep	LARSLGVASIRVVETILGKTCMGLEL	PNPKRQMI	RSEIFNSPEFAESKSKLT	LALGQDI		
10	orf58-1	LARSLGVASIRVVETIPGKTCMGLEL	PNPKRQMI	RSEIFNSPEFAESKSKLT	LALGQDI		
		610	620	630	640	650	660
	orf58a.pep	TGQPVVTDLGKAPHLLVAGTTGSGK	SVGVNAMILSMLFKAAPEDVR	MIMIDPKM	LELSIY		
15	orf58-1	TGQPVVTDLGKAPHLLVAGTTGSGK	SVGVNAMILSMLFKAAPEDVR	MIMIDPKM	LELSIY		
		670	680	690	700	710	720
	orf58a.pep	EGIPHLLAPVVTDMKLAANALNWC	VNEMEKRYRLMSFMGVRNLAGX	NQKIAEAAARGEKI			
20	orf58-1	EGIPHLLAPVVTDMKLAANALNWC	VNEMEKRYRLMSFMGVRNLAGF	NQKIAEAAARGEKI			
		730	740	750	760	770	780
	orf58a.pep	GNPFSLTDPNPEPLXKLPIVVVVDE	FADLMMTAGKKIEELIARLAQ	KARAAGIHLILAT			
25	orf58-1	GNPFSLTDPDPEPLEKLPFIVVVDE	FADLMMTAGKKIEELIARLAQ	KARAAGIHLILAT			
		790	800	810	820	830	840
	orf58a.pep	QRPSVDVITGLIKANIPTRIAFQV	SSKIDSRTILDQMGAEENLLGQ	GDMFLP	PGTAYPQR		
30	orf58-1	QRPSVDVITGLIKANIPTRIAFQV	SSKIDSRTILDQMGAEENLLGQ	GDMFLP	PGTAYPQR		
		850	860	870	880	890	900
	orf58a.pep	VHGAFASDEEVHRVVEYLKQFGE	PDYVDDXLSSGMSDDLGLIS	RSRGDGETDPMYDEAVSV			
35	orf58-1	VHGAFASDEEVHRVVEYLKQFGE	PDYVDDILSGGSEELPGIGR	SGDDETDPMYDEAVSV			
		910	920	930	940	950	960
	orf58a.pep	VLKTRKASISGVQRALRIGYNRAAR	LIDQMEAGIVSAPEHNGNRTIL	VPLDNAX			
40	orf58-1	VLKTRKASISGVQRALRIGYNRAAR	LIDQMEAGIVSAPEHNGNRTIL	VPLDNAX			
		970	980	990	1000	1010	

# 50 Homology with a predicted ORF from *N.gonorrhoeae*

ORF58 shows complete identity over a 9aa overlap with a predicted ORF (ORF58ng) from *N. gonorrhoeae*:

	orf58.pep	ALMLFHAVKTA	VYWL	LVGV	VRF	CRNYLA	HESE	PD	RP	VPP	103
55	orf58ng										

The ORF58ng nucleotide sequence <SEQ ID 493> is predicted to encode a protein having partial amino acid sequence <SEQ ID 494>:

	1	..	SEPDRPVPPA	SANRADVPTA	SDGYSDSGNG	TEEAETAAE	AAEEEEADTE	
60	51		DIATAVIDNR	RIPFDRSIAE	GLMQSEKTS	PVRPVFKEIT	LEEATRALSS	
	101		AALRETKKRY	IDAFEKNGTA	VPKVRVSDTP	MEGLQIIGLD	DPVLQRTYSR	
	151		MFDADKEAFS	ESADYGFEPY	FEKQHPSAFS	AVKAENARNA	PFRRHAGQEK	
	201		GQAEAKSPDV	SQGQSVSDGT	AVRDARRRVS	VNLKEPNKAT	VSAEARISRL	
	251		IPESRTVVGK	RDVEMPSETE	NVFTETVSSV	GYGGPVYDEA	ADIHIEEPAA	
	301		PDWVVVEPPE	VPEVAVPEID	ILPPPPVSEI	YNRTYEPPAG	FEQAQRSRIA	

-291-

351 ETDHLAADVL NGGWQEETAA IADDGSEGAA ERSSGQYLSE TEAFGHDSQA  
 401 VCPFEDVPSE RPSCRVSDE ADEGAFQSEE TGAVSEHLPT TDLLLPPLFN  
 451 PEATQTEEL LENSITIEEK LAEFKVKVKV VDSYSGPVIT RYEIEPDVGV  
 501 RGNVSLNLEK DLARSLGVAS IRVVETIPGK TCMGLELPNP KRQMIRLSEI  
 551 FNSPEFAESK SKLTALGQD ITGQPVVTDL GKAPHLLVAG TTGSGKSVGV  
 601 NAMILSMLFK AAPEDVRMIM IDPKMLELSI YEGITHLLAP VVTDMLKLAN  
 651 ALNWCVNEME KRYRLMSFMG VRNLAGFNQK IAEAAARGEK IGNPFSLTDP  
 701 DPEPLEKLPF IVVVVDEFAD LMMTAGKKIE ELIARLAQKA RAAGIHLILA  
 751 TQRPSVDVIT GLIKANIPTR IAFQVSSKID SRTILDQMG A ENLLGQGDML  
 801 FLPPGTAYPQ RVHGAFASDE EVHRVVEYLK QFGEPTYDD ILSGGGSEEL  
 851 PGIGRSGDGE TDPYDEAVS VVLKTRKASI SGVQRALRIG YNRAARLIDQ  
 901 MEAEGIVSAP EHNGNRTILV PLDNA\*

This partial gonococcal sequence contains a predicted transmembrane region and a predicted  
 ATP/GTP-binding site motif A (P-loop; double underlined). Furthermore, it has a domain  
 15 homologous to the FTSK cell division protein of *E. coli*. Alignment of ORF58ng and FtsK  
 (accession number p46889) show a 65 % amino acid identity in 459 overlap:

ORF58ng: 467 IEKLAEFKVKVKVDSYSGPVITRYEIEPDVGVGRNSVLNLEKDLARSLGVASIRVVET 526  
 +E +LA+F++K VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVVE  
 FtsK: 868 VEARLADFRKADVVNYSPPGVITRFEFLNAPGVKAARISNLSRDLARSLSTVAVRVVEV 927  
 ORF58ng: 527 IPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDITGQPVVTDLGKAPHL 586  
 IPGK +GLELPN KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVV DL K PHL  
 FtsK: 928 IPGKPYVGLELPNKKRQTVYLRVLDNAKFRDNPSPLTVVLGKDIAGEPVVADLAKMPHL 987  
 ORF58ng: 587 LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDMMK 646  
 LVAGTTGSGKSVGVNAMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDMMK  
 FtsK: 988 LVAGTTGSGKSVGVNAMILSMLYKAQPEDVRFIMIDPKMLELSVYEGIPHLLETVVTDMMK 1047  
 ORF58ng: 647 LAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKIGNPFSLTDPDDPEP-- 704  
 AANAL WCVNEME+RY+LMS +GVRNLAG+N+KIAEA I +P+ D +  
 FtsK: 1048 DAANALRWCVNEMERRYKLMSALGVRNLAGYNEKIAEADRMMPPIPDYKPGDSMDAQH 1107  
 ORF58ng: 705 --LEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPSVDVITGL 762  
 L+K P+IVV+VDEFADLMMT GKK+EELIARLAQKARAAGIHL+LATQRPSVDVITGL  
 FtsK: 1108 PVLKKEPYIVVLVDEFADLMMTVGKKVEELIARLAQKARAAGIHLVLATQRPSVDVITGL 1167  
 ORF58ng: 763 IKANIPTRIAFQVSSKIDSRITLDQMGAEENLLGQGDMLFLPPGTAYPQVRHGAFASDEEV 822  
 IKANIPTRIAF VSSKIDSRITLDQ GAE+LLG GDML+ P + P RVHGAF D+EV  
 FtsK: 1168 IKANIPTRIAFTVSSKIDSRITLDQAGAESLLGMGDMLYSGPNSTLPVRVHGAFVRDQEV 1227  
 ORF58ng: 823 HRVVEYLKQFGEPTYDDILSGGGSEELPGIGRSGDGETDPYDEAVSVVLKTRKASISG 882  
 H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG  
 FtsK: 1228 HAVVQDWKARGRPQYVDGITSDESEGGAG-GFDGAEELDPLFDQAVQFVTEKRKASISG 1286  
 ORF58ng: 883 VQRALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVP 921  
 VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P  
 FtsK: 1287 VQRQFRIGYNRAARIIEQMEAQGVSEQQHNGNREVLAP 1325

Further work on ORF58ng revealed the complete gonococcal DNA sequence to be <SEQ ID 495>:

1 ATGTTTTGGA TAGTTTTGAT CGTTATtgtg TTGCTTGCGC TTGCCGGCCT  
 50 51 GTTTTTTGTG CGCGCACAAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG  
 101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA  
 151 GACGGTATGC CCGATTTTCC CGAGTTTTCG CTGATGCTTT TCCATGCCGT  
 201 CAAAACGGCA GTGTATTGGC TGTGTGTCGG TGTGTCCTCGT TTCTGCGCAA  
 251 ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT  
 55 301 GCAAACCGTG CGGATGTTCC GACCGCATCC GACGGGTATT CAGACAGTGG  
 351 AAACGGGACG GAAGAAGCGG AAACGGAAGC AGCAGAAGCT GCGGAGGAAG  
 401 AGGCTGCCgA TACgGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC  
 451 ATCCcatTCG ACCGAGTAT TGCTGAAGGG TTGATGCAGT CTGAAAGCAA  
 501 AACTTCGCCC GTCCGTCGGG TTTTAAAGGA AATCACTTTG GAAGAAGCAA  
 60 551 CGCGTGCTTT AAGCAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC  
 601 GATGCATTG AGAAAAACGG AACAGCCGTC CCCAAAGTAC GCGTGTCCGA  
 651 TACCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC  
 701 AACGCACGTA TTCCCGTATG TTTGATGCGG ACAAAGAAGC GTTTTCCGAG  
 751 TCTGCGGATT ACGGATTGTA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC

	801	CTTTTCTGCA	GTCAAAGCCG	AAAATGCACG	GAATGCGCCG	TTCCGCCGTC
	851	ATGCAGGGCA	GGAGAAAGGG	CAGGCGGAGG	CAAAATCCCC	GGATGTTTCC
	901	CAAGGGCAGT	CCGTTTTCAGA	CGGCACAGCC	GTCCGCGATG	CCCCGCCGCG
5	951	CGTTTCCGTC	AATTTGAAAG	AACCGAACAA	GGCAACGGTT	TCTGCGGAGG
	1001	CGCGGATTTT	GGCGCTGATT	CCGGAAGATC	GGACGGTTGT	CGGGAACCGG
	1051	GATGTCGAAA	TGCCGTCTGA	AACCGAAAAT	GTTTTACCGG	AAACCGTTTC
	1101	GTCTGTGGGA	TACGGCGGTC	CGGTTTATGA	TGAAGCTGCC	GATATCCATA
	1151	TTGAAGAGCC	TGCCGCGCCC	GATGCTTGGG	TGGTCGAACC	ACCCGAAGTG
	1201	CCGGAGGTAG	CCGTACCCGA	AATCGATATT	CTGCCGCCGC	CTCCCGTATC
10	1251	GGAAATCTAC	AACCGTACCT	ATGAGCCGCC	GGCAGGATTC	GAGCAGGCGC
	1301	AACGCAGCCG	CATTGCCGAA	ACCGACCATC	TTGCCGCTGA	TGTTTTGAAT
	1351	GGAGGTTGGC	AGGAGGAAAC	CGCCGCTATT	GCAGATGACG	GCAGTGAGGG
	1401	TGCGGCAGAG	CGGTCAAGCG	GGCAATATCT	GTCGGAAACC	GAAGCGTTCG
	1451	GGCATGACAG	TCAGGCGGTT	TGTCCGTTTG	AAGATGTGCC	GTCTGAACGC
15	1501	CCGTCTGCC	GGGTATCGGA	TACGGAAGCG	GATGAAGGGG	CGTTCCATC
	1551	GGAAGAGACC	GGTGCGGTAT	CCGAACACCT	GCCGACAACC	GACCTGCTTC
	1601	TGCCTCCGCT	GTTCAATCCC	GAGGCGACGC	AAACCGAAGA	AGAACTGTTG
	1651	GAAACAGCA	TCACCATCGA	AGAAAAATG	GCGGAGTTCA	AAGTCAAGGT
20	1701	CAAGGTTGTC	GATTCTTATT	CCGCCCCCGT	GATTACGCGT	TATGAAATCG
	1751	AACCCGATGT	CGCGGTGCGC	GGCAATTCCG	TTCTGAATTT	GGAAAAAGAC
	1801	TTGGCGCGTT	CGCTCGGCGT	GGCTTCCATC	CGCGTTGTCG	AAACCATCCC
	1851	CGGCAAAACC	TGCATGGGTT	TGGAACCTCC	GAACCCGAAA	CGCCAAATGA
	1901	TACGCCTGAG	CGAAATTTTC	AATTCGCCCG	AGTTTGCCGA	ATCCAAATCC
25	1951	AAGCTGACGC	TCGCGCTCGG	TCAGGACATT	ACCGGACAGC	CCGTCGTAAC
	2001	CGACTTGGGC	AAAGCACC GC	ATTTGCTGGT	TGCCGGCAGC	ACCGGTTCCG
	2051	GCAAATCGGT	GGGTGTCAAC	GCGATGATTC	TGTCTATGCT	TTTCAAAGCC
	2101	GCGCCGGAAG	ACGTGCGTAT	GATTATGATC	GATCCGAAAA	TGCTGGAATT
	2151	GAGCATTTAC	GAAGGCATCA	CGCACCTGCT	CGCCCCTGTC	GTTACCGATA
30	2201	TGAAGCTGGC	GGCAAACGCG	CTGAACGGT	GTGTTAACGA	AATGGAAAAA
	2251	CGCTACCGCC	TGATGAGCTT	TATGGGCGTG	CGCAATCTTG	CGGGCTTCAA
	2301	CCAAAAAATC	GCCGAAGCCG	CAGCAAGGGG	AGAAAAAATC	GGCAATCCGT
	2351	TCAGCCTCAC	GCCCGACGAT	CCCGAACCTT	TGGAAAAACT	GCCGTTTATC
	2401	GTGGTCGTGG	TCGATGAGTT	TGCCGATTTG	ATGATGACGG	CAGGCAAGAA
35	2451	AATCGAAGAA	CTGATTGCGC	GCCTCGCCCA	AAAAGCCCCG	GCGGCAGGCA
	2501	TCCACCTTAT	CCTTGCCACA	CAACGCCCCA	GCGTCGATGT	CATCACGGGT
	2551	CTGATTAAGG	CGAACATCCC	GACGCGTATC	GCGTTCCAAG	TGTCCAGCAA
	2601	AATCGACAGC	CGCACGATTC	TCGACCAAAT	GGGCGCGGAA	AACCTGCTCG
	2651	GTCAAGGCGA	TATGCTGTTT	CTGCCGCCGG	GTA CTGCTA	TCCGACGCGC
40	2701	GTTACGCGCG	CGTTTGCCCT	GGATGAAGAG	GTGCACCGCG	TGGTCGAATA
	2751	TCTGAAGCAG	TTTGCGCAGC	CGGACTATGT	TGACGATATT	TTGAGCGCGC
	2801	GCGGCAGCGA	AGAGCTGCCC	GGCATCGGGC	GCAGCGGCGA	CGGCGAAACC
	2851	GATCCGATGT	ACGACGAGGC	CGTATCCGTT	GTCCTGAAAA	CGCGCAAAGC
	2901	CAGCATTTTC	GGCGTACAGC	GCGCCTTGCG	CATCGGTAC	AACCGCGCCG
45	2951	CGCGTCTGAT	TGACCAAATG	GAAGCGGAAG	GCATTGTGTC	CGCACCGGAA
	3001	CACAACGGCA	ACCGTACGAT	TCTCGTCCCC	TTGGACAATG	CTTGA

This corresponds to the amino acid sequence <SEQ ID 496; ORF58ng-1>:

	1	MFWIVLIVIV	LLALAGLFFV	RAQSEREWMR	EVSAWQEKKG	EKQAE LPEIK
	51	DGMPDFPEFS	LMLFHAVKTA	VYWLFGVVVR	FCRNYLAHES	EPDRPVPPAS
50	101	ANRADVPTAS	DGYSDSNGT	EEAETEAAEA	AEEEEADTED	IATAVIDNRR
	151	IPFDRSIAEG	LMQSESKTSP	VRPVFKEITL	EEATRALSSA	ALRETKKRYI
	201	DAFEKNGTAV	PKVRVSDTPM	EGLQIIGLDD	PVLQRTYSRM	FDADKEAFSE
	251	SADYGFEPYF	EKQHPSAFSA	VKAENARNAP	FRRHAGQEK	QAEAKSPDVS
	301	QQQSVSDGTA	VRDARRRVSV	NLKEPNKATV	SAEARISRLI	PESRTVVGKR
55	351	DVEMPSETEN	VETETVSSVG	YGGPVYDEAA	DIHIEEPAAP	DAWVVEPPEV
	401	PEVAVPEIDI	LPPPPVSEIY	NRTYEPPAGF	EQAQRSRIAE	TDHLAADVLN
	451	GGWQEETA AI	ADDGSEGAEE	RSSGQYLSET	EAFGHDSQAV	CPFEDVP SER
	501	PSCRVS DTEA	DEGA FQSEET	GAVSEHLPTT	DLLLPLPLFN	EATQT EEEELL
	551	ENSITIEEKL	AEFKVKVKVV	DSYSGPVITR	YEIEPDVGVR	GNSVLNLEKD
60	601	LARSLGVASI	RVVETIPGKT	CMGLELPNPK	RQMIRLSEIF	NSPEFAESKS
	651	KLTLALGQDI	TGQPVVTDLG	KAPHLLVAGT	TGSGKSVGVN	AMILSMLFKA
	701	APEDVRMIMI	DPKMLELSIY	EGITHLLAPV	VTDMKLAANA	LNWCVNEMEK
	751	RYRLMSFMGV	RNLAGFNQKI	AEAAARGEKI	GNPFSLTPDD	PEPLEKLPFI
	801	VVVVDEFADL	MMTAGKKIEE	LIARLAQKAR	AAGIHLILAT	QRPSVDVITG
65	851	LIKANIPTRI	AFQVSSKIDS	RTILDQMGAE	NLLGQGDMLF	LPPGTAYPQR
	901	VHGAFASDEE	VHRVVEYLKQ	FGEPDYVDDI	LSGGGSEELP	GIGRSGDGET
	951	DPMYDEAVSV	VLKTRKASIS	GVQRALRIGY	NRAARLIDQM	EAEGIVSAFE
	1001	HNGNRTILVP	LDNA*			

ORF58ng-1 and ORF58-1 show 97.2% identity in 1014 aa overlap:

-293-

		10	20	30	40	50	60
	orf58-1.pep	MFWIVLIVILLALAGLFFVRAQSEREWREVS AWQEKKGKQAE LPEIKDGM PDPFPELA					
	orf58ng-1	MEWIVLIVIVLLALAGLFFVRAQSEREWREVS AWQEKKGKQAE LPEIKDGM PDPFPEFS					
5		10	20	30	40	50	60
	orf58-1.pep	70	80	90	100	110	120
	orf58ng-1	LMLFHAVKTAVYWLFGVVRFCRNYLAHESEPD RPVP PASANRADVPTASDGYSDSGNGT					
10		70	80	90	100	110	120
	orf58-1.pep	130	140	150	160	170	180
	orf58ng-1	EEAETEEAEAEAEAAADTEDIATAVIDNRRIPFDRSIAEGLMPSESEISPV RPVFKEITL					
15		130	140	150	160	170	180
	orf58-1.pep	190	200	210	220	230	240
	orf58ng-1	EEATRALNSAALRETKKRYIDAFEKNETAVPKVRVSDTPMEGLQIIGLDDPVLQRTYSHM					
20		190	200	210	220	230	240
	orf58-1.pep	250	260	270	280	290	300
	orf58ng-1	FDADKEAFSEADYGFEPYFEKQHPSAFSAVKAENARNAPFRRHAGQGKGQAEAKSPDVS					
25		250	260	270	280	290	300
	orf58-1.pep	310	320	330	340	350	360
	orf58ng-1	QGQSVSDGTAVRDARRRVSVNLKEPNKATVSAEARISRLIPESQTVVGKRDVEMPSETEN					
30		310	320	330	340	350	360
	orf58-1.pep	370	380	390	400	410	420
	orf58ng-1	VFTETVSSVGYPVYDETAADIHIEEPAAPDAWVVEPEVPEVKVPM TADIIQPPPPVSEIY					
35		370	380	390	400	410	420
	orf58-1.pep	430	440	450	460	470	480
	orf58ng-1	NRTYEPPSGFEQVQRSRIAETDHLADDVLNGGWQEETA AIADDGSEGA AERSSSGQYLSET					
40		430	440	450	460	470	480
	orf58-1.pep	490	500	510	520	530	540
	orf58ng-1	EAFGHDSQAVCPFENVPSERPSCRVSDTEAD EGAFFPSEETGAVSEHLPTDLLLPPLFNP					
45		490	500	510	520	530	540
	orf58-1.pep	550	560	570	580	590	600
	orf58ng-1	EATQTEEELENSITIEEKLAEFKVKVVD SYSGPVITRYEIEPDVGVRGNSVLNLEKD					
50		550	560	570	580	590	600
	orf58-1.pep	610	620	630	640	650	660
	orf58ng-1	LARSLGVASIRVVETIPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLT LALGQDI					
55		610	620	630	640	650	660
	orf58-1.pep	670	680	690	700	710	720
	orf58ng-1	TGQPVVTDLGKAPHLLVAGTTGSGKSVGNAMILSMLFKAAPEDVRMIMIDPKMLELSIY					
60		670	680	690	700	710	720
	orf58-1.pep						
	orf58ng-1						
65		670	680	690	700	710	720
	orf58-1.pep						
	orf58ng-1						
70		670	680	690	700	710	720
	orf58-1.pep						
	orf58ng-1						

-294-

		730	740	750	760	770	780
	orf58-1.pep	EGIPHLLAPVVTDMKLAANALNWCNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI					
5	orf58ng-1	EGITHLLAPVVTDMKLAANALNWCNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI					
		730	740	750	760	770	780
	orf58-1.pep	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
10	orf58ng-1	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
		790	800	810	820	830	840
	orf58-1.pep	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLLPGTAYPQR					
15	orf58ng-1	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLLPGTAYPQR					
		850	860	870	880	890	900
	orf58-1.pep	VHGAFASDEEVHRVVEYLKQFGEVDYDDILSGGGSEELPGIGRSGDDETDPMYDEAVSV					
20	orf58ng-1	VHGAFASDEEVHRVVEYLKQFGEVDYDDILSGGGSEELPGIGRSGDDETDPMYDEAVSV					
		910	920	930	940	950	960
	orf58-1.pep	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVPLDNAX					
25	orf58ng-1	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVPLDNAX					
		970	980	990	1000	1010	

30 Furthermore, ORF58ng-1 shows significant homology to the *E.coli* protein FtsK:

35 sp|P46889|FTSK\_ECOLI\_CELL\_DIVISION\_PROTEIN\_FTSK >gi|1651412|gnl|PID|d1015290 (D1 division protein FtsK [Escherichia coli]) >gi|1651418|gnl|PID|d1015296 (D90727) Cell division protein FtsK [Escherichia coli] >gi|1787117 (AE000191) cell division protein FtsK [Escherichia coli] Length = 1329  
 Score = 576 bits (1469), Expect = e-163  
 Identities = 301/459 (65%), Positives = 353/459 (76%), Gaps = 5/459 (1%)

Query: 556 IEEKLAEFKVKVVDVSYSGPVITRYEIEPDVGVGRNSVLNLEKDLARSLGVASIRVVET 615  
 +E +LA+F++K VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVVE  
 40 Sbjct: 868 VEARLADFRIKADVNNYSPGPVITRFELNLAPGVKAARISNLSRDLARSLSTVAVRVVEV 927

Query: 616 IPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDITGQPVVTDLGKAPHL 675  
 IPGK +GLELPN KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVV DL K PHL  
 45 Sbjct: 928 IPGKPYVGLELPNKKRQTVYLREVLNDAKFRDNPSPLTVVLGKDIAGEPVVADLAKMPHL 987

Query: 676 LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDMK 735  
 LVAGTTGSGKSVGVNAMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDMK  
 50 Sbjct: 988 LVAGTTGSGKSVGVNAMILSMLYKAQPEDVRFIMIDPKMLELSVYEGIPHLLTEVVTDMK 1047

Query: 736 LAANALNWCNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKIGNPFSLTPDDPEP-- 793  
 AANAL WCVNEME+RY+LMS +GVRNLAG+N+KIAEA I +P+ D +  
 55 Sbjct: 1048 DAANALRWCVNEMERRYLKMSALGVRNLAGYNEKIAEADRMMPRIPIPDYWKPGDSMDAQH 1107

Query: 794 --LEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPSVDVITGL 851  
 L+K P+IVV+VDEFADLMMT GKK+EELIARLAQKARAAGIHL+LATQRPSVDVITGL  
 60 Sbjct: 1108 PVLKKEPYIVVLVDEFADLMMTVGKKVEELIARLAQKARAAGIHLVLATQRPSVDVITGL 1167

Query: 852 IKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLLPGTAYPQRVHGAFASDEEV 911  
 IKANIPTRIAF VSSKIDSRTILDQ GAE+LLG GDML+ P + P RVHGAF D+EV  
 65 Sbjct: 1168 IKANIPTRIAFTVSSKIDSRTILDQAGAESLLGMGDMLYSGPNSTLPVRVHGAFVRDQEV 1227

Query: 912 HRVVEYLKQFGEVDYDDILSGGGSEELPGIGRSGDGETDPMYDEAVSVVLKTRKASISG 971  
 H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG  
 70 Sbjct: 1228 HAVVQDWKARGRPQYVDGITSDESEGGAG-GFDGAEELDPLFDQAVQFVTEKRKASISG 1286

Query: 972 VQRALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVP 1010  
 VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P  
 75 Sbjct: 1287 VQRQFRIGYNRAARIIEQMEAQGIVSEQGHNGNREVLAP 1325

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 59

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 497>:

```

5      1  ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
      51  CATTTTCGTC GTCCTCTTGG CGGTATTGGT CTCCACGCAG GCAATCAACC
      101 TGCTCGGCCG TGCCGCCGAC GGGC..GTGA TCGCCATCGA TGCCGTGTTG
      151 GCATTGGTCG GCTTCTGGGT C.....
//
10     901 .....A TTGCCATCGG TTTGTTTTTA ATTTACCAA ACGGGCTGAC
      951 CCTGCTTTTT GAAGCCGTGG AAGACGGCAA AATCCATTTT TGGCTCGGAC
      1001 TGCTGCCATAT GCACATTATC ATGTTTGTCC TTGCACTCAT CCTGTGCGC
      1051 TGCCGCAGTA TGCCAGCCA GCCCTTCTGG CAGGCGGTTG GCAAAAGTCT
      1101 GACATTGAAA GGCGGAAAT GA

```

15 This corresponds to the amino acid sequence <SEQ ID 498; ORF101>:

```

      1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GXVIAIDAVL
      51  ALVGFWV... .....
//
      301 ...IAIGLFL IYQNLTLLE EAVEDGKIHF WLGLLPMHII MFVLALILLR
      351 VRSMPSQPFW QAVGKSLTLK GKK*
20

```

Further work revealed the complete nucleotide sequence <SEQ ID 499>:

```

      1  ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
      51  CATTTTCGTC GTCCTCTTGG CGGTATTGGT CTCCACGCAG GCAATCAACC
      101 TGCTCGGCCG TGCCGCCGAC GGGCGTGTG CCGTATGATG CGTGTGCGA
25     151 TTGGTCGGCT TCTGGGTCAT CCGTATGACG CCGCTTTTGC TGGTGTGAC
      201 CGCATTATATC AGTACGTTGA CCGTGTGAC CCGCTACTGG CCGACAGCG
      251 AAATGTCGGT CTGGCTATCC TGCGGATTGG CATTGAAACA ATGGATACGC
      301 CCGGTGATGC AGTTTGCCGT GCCGTTTGCC GTTTTGGTTG CCGTCATGCA
      351 GCTTTGGGTG ATACCGTGGG CAGAGCTACG CAGCCGCGAA TACGCTGAAA
30     401 TCCTGAAGCA GAAGCAGGAA TTGTCTTTGG TGGAGGCAGG CGAGTTCAAC
      451 AGTTTGGGCA AGCGCAACGG CAGGGTTTAT TTTGTCGAAA CCTTCGATAC
      501 CGAATCCGGC ATCATGAAAA ACCTGTTCCG GCGCGAACAG GACAAAAACG
      551 GCGGCGACAA CATCATCTTC GCCAAAGAAG GTAACCTCTC GCTGAACGAC
      601 AACAAACGCA CGCTCGAATT GCGCCACGGC TACCGTTACA GCGGCACGCC
35     651 CCGACGCGCC GACTACAATC AGGTTTCCTT CCAAAACTC AACCTGATTA
      701 TCAGCACCAC GCCCAAATC ATCGACCCCG TTTCCACCG CCGTACCATT
      751 CCGACGCCCC AACTGATTGG CAGCAGCAAC CCGCAACATC AGGCGGAATT
      801 GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTACTC TGCTGCTTG
      851 CCGTGCCGCT TTCCTATTTC AACCCGCGCA GCGGACATAC CTACAATATC
40     901 TTGATTGCCA TCGGTTTGTG TTTAATTTAC CAAAACGGGC TGACCCTGCT
      951 TTTTGAAGCC GTGGAAGACG GCAAAATCCA TTTTGGCTC GGACTGCTGC
      1001 CTATGCACAT TATCATGTTT GCCGTTGCAC TCATCCTGTT GCGGTCGCGC
      1051 AGTATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT
      1101 GAAAGCGGA AAATGA

```

45 This corresponds to the amino acid sequence <SEQ ID 500; ORF101-1>:

```

      1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
      51  LVGFVIGMT PLLLVLTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR
      101 PVMQFAVPFA VLVAVMQLWV IPWAE LRSRE YAEILKQKQE LSLVEAGEFN
50     151 SLGKRNRRVY FVETFDTESG IMKNLFLREQ DKNNGDNIIF AKEGNFSLND
      201 NKRTLELRHG YRYSPTPGRA DYNQVSFQKL NLIISTPKL IDPVSHRRTI
      251 PTAQLIGSSN PQHQAEMLWR ISLTVSVLLL CLLAVPLSYF NPRSGHTYNI
      301 LIAIGLFLIY QNGLTLLFEA VEDGKIHFVW GLLPMHIIMF AVALILLRVR
      351 SMPSQPFWQA VGKSLTLKGG K*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF101 shows 91.2% identity over a 57aa overlap and 95.7% identity over a 69aa overlap with an ORF (ORF101a) from strain A of *N. meningitidis*:

5	orf101.pep	MIYQRNLIKELSFTAVGIFVVLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFVWX	10	20	30	40	50
	orf101a	MIYQRNLIKELSFTAVGIFVVLAVLVSTQAINLLGXAADXR- AIDAVLALVGFVWXXM	10	20	30	40	50
10	orf101.pep	.....IAIGLFLIYQNGLTLLFEAVEDGKIHFWLGL	//				
	orf101a	LTVSVLLLCLLAVPLSYFNPRSGHTYNILXAIGLFLIYQNGLTLLFEAVEDGKIHFWLGL	280	290	300	310	320
15	orf101.pep	LPMHIIMFVLALILLRVRSMPSQPFQAVGKSLTLKGGKX	120	130	140	150	
	orf101a	LPMHIIMFVIAIVLLRVRSMPSQPFQAVGKSLTLKGGKX	340	350	360	370	

The complete length ORF101a nucleotide sequence <SEQ ID 501> is:

1	ATGATTTATC	AAAGAAACCT	CATCAAAGAA	CTCTCTTTTA	CCGCCGTCGG
51	CATTTTCGTC	GTCCTCTTGG	CGGTATTGGT	CTCCACGCAG	GCAATCAACC
101	TGCTCGGCCN	TGCCGCCGAC	NGGCGTNTCG	CCATCGATGC	CGTGTGGCA
151	TTGGTCGGCT	TCTGGGTCNN	NNGNATGACG	CCGCTTTTGC	TNGTGTGAC
201	CGCATTTATC	AGTACGTTGA	CCGTGTGAC	CCGCTACTGG	CGNGACAGCG
251	AAATGTCGGT	CTGGNTATCC	TGCGGATTGG	CATTGAAACA	ATGGATACGC
301	CCGGTGATGC	AGTTTGCCGT	GCCGTTTGCC	GTTTGGGTG	CCGTCATGCA
351	GCTTTGGGTG	ATACCGTGGG	CAGAGCTACG	CAGCCGCGAA	TACGCTGAAA
401	TCCTGAAGCA	GAAGCAGGAA	TTGTCTTTGG	TGGAGGCAGG	CGGGTTCAAC
451	AGTTTGGGCA	AGCGCAACGG	CAGGGTTTAT	TTTGTGCGAA	CCTTCGATAC
501	CGAATCCGGC	ATCATGAAAA	ACCTGTTTCT	GCGCGAACAG	GACAAAAACG
551	GCGGCGACAA	CATCATCTTC	NCCAAAGAAA	GTAACCTCTC	GCTGAACGAC
601	AACAAACGCA	CGCTCGAATT	GCGCCACGGC	TACCGTTACA	GCGGCACGCC
651	CGGACGCGCC	GACTACAATC	AGGTTTCCTT	CCNAAAACCTC	AACCTGATTA
701	TCAGCACCAC	GCCCAAATC	ATCGACCCCG	TTTCCCACCG	CCGTACNATN
751	CCNACNGCCC	AACTGATTGG	CAGCAGCAAC	CCGCAACATC	ANGCGGAATT
801	GATGTGGCGC	ATCTCGCTGA	CCGTACGCGT	CCTCCTACTC	TGCTGCTTG
851	CCGTGCCGCT	TTCCTATTTC	AACCCGCGCA	GCGGACATAC	CTACAATATC
901	TGANTGCCA	TCGGTTTGT	TTTAATTAC	CAAAACGGGG	TGACCTGCT
951	TTTTGAAGCC	GTGGAAGACG	GCAAAATCCA	TTTTTGGCTC	GGACTGCTGC
1001	CTATGCACAT	CATCATGTTC	GTCATCGCAA	TCGTAATTCT	GCGCGTCCGC
1051	AGCATGCCCA	GCCAGCCCTT	CTGGCAGGCG	GTTGGCAAAA	GTCTGACATT
1101	GAAAGCGCGA	AAATGA			

45 This encodes a protein having amino acid sequence <SEQ ID 502>:

1	MIYQRNLIKE	LSFTAVGIFV	VLLAVLVSTQ	AINLLGXAAD	XRXAIDAVLA
51	LVGFVWXXMT	PLLLVLTAFI	STLTVLTRYW	RDSEMSVWXS	CGLALKQWIR
101	PVMQFAVPFA	VLVAVMQLWV	IPWAEIARSRE	YAEILKQKQE	LSLVEAGGFN
151	SLGKRNRVY	FVETFDTESG	IMKNLFLREQ	DKNGGDNIIIF	XKESNFSLND
201	NKRTLELRHG	YRYSPTPGRA	DYNQVSFXKL	NLIISTTPKL	IDPVSHRRTX
251	PTAQLIGSSN	PQHXAELMWR	ISLTVSVLLL	CLLAVPLSYF	NPRSGHTYNI
301	LXAIGLFLIY	QNGLTLLFEA	VEDGKIHFWL	GLLPMHIIMF	VIAIVLLRVR
351	SMPSQPFQQA	VGKSLTLKGG	K*		

ORF101a and ORF101-1 show 95.4% identity in 371 aa overlap:

55	orf101a.pep	MIYQRNLIKELSFTAVGIFVVLAVLVSTQAINLLGXAADXRXAIDAVLALVGFVWXXMT	60
	orf101-1	MIYQRNLIKELSFTAVGIFVVLAVLVSTQAINLLGRAADGRVAIDAVLALVGFVWIGMT	60
60	orf101a.pep	PLLLVLTAFIISTLTVLTRYWRDSEMSVWXSGLALKQWIRPVMQFAVPFAVLVAVMQLWV	120
	orf101-1	PLLLVLTAFIISTLTVLTRYWRDSEMSVWLSGLALKQWIRPVMQFAVPFAVLVAVMQLWV	120



	orf101a.pep	IPWAE LRSREYAEILKQKQELSLVEAGGFNSLGKRNGRVYFVETFDTESGIMKNLFLREQ	180
	orf101-1	IPWAE LRSREYAEILKQKQELSLVEAGEFNSLGKRNGRVYFVETFDTESGIMKNLFLREQ	180
5	orf101a.pep	DKNGGDNIIIFXKESNFS LNDNKRTLELRHGYRSGTPGRADYNQVSFXKLNLIISTTPKL	240
	orf101-1	DKNGGDNIIIFAKEGNFS LNDNKRTLELRHGYRSGTPGRADYNQVSFQKLNLIISTTPKL	240
10	orf101a.pep	IDPVSHRRTXPTAQLIGSSNPQHXAELMWRLSLTVSVLLLCLLAVPLSYFNPRSGHTYNI	300
	orf101-1	IDPVSHRRTIPTAQLIGSSNPQHQAELMWRLSLTVSVLLLCLLAVPLSYFNPRSGHTYNI	300
15	orf101a.pep	LXAIGLFLIYQNGLTLLFEAVEDGKIHFWLGLLPMHIIMFVIAIVLLRVRSMPSQPFWQA	360
	orf101-1	LIAIGLFLIYQNGLTLLFEAVEDGKIHFWLGLLPMHIIMFAVALILLRVRSMPSQPFWQA	360
	orf101a.pep	VGKSLTLKGGK	371
20	orf101-1	VGKSLTLKGGK	371

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF101 shows 96.5 % identity in 57aa overlap at the N-terminal domain and 95.1% identity in 61aa overlap at the C-terminal domain, respectively, with a predicted ORF (ORF101ng) from *N.*

*gonorrhoeae*:

	orf101.pep	MIYQRNLIKELSFTAVGIFVVLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFVW	57
	orf101ng	MIYQRNLIKELSFTAVGIFVVLAVLVSTQAINLLGRAADGRV-AIDAVLALVGFVWIGM	59
30		//	
	orf101.pep	IAIGLFLIYQNGLTLLFEAVEDGKIHFWLG	333
35	orf101ng	SLTVSVLLLCLLAVPLSYFNPRSGHTYNILIAIGLFLIYQNGLTLLFEAVEDGKIHFWLG	331
	orf101.pep	LLPMHIIMFVLALILLRVRSMPSQPFWQAVGKSLTLKGGK	373
	orf101ng	LLPMHIIMFVIAIVLLRVRSMPSQPFWQAVG	362

The ORF101ng nucleotide sequence <SEQ ID 503> is predicted to encode a protein having partial amino acid sequence <SEQ ID 504>:

	1	MIYORNIKE	LSFTAVGIFV	VLLAVLVSTQ	AINLLGRAAD	GRVAIDAVLA
	51	LVGFVWIGMT	PLLLVLTAFI	STLTVLTRYW	RDSEMSVWLS	CGLALKQWIR
	101	PVMQFAVPFA	ILIAVMQLWV	IPWAE LRSRE	YAEILKQKQE	LSLVEAGEFN
45	151	NLGKRNGRVY	FVETFDTESG	IMKNLFLREQ	DKNGGDNIIIF	AKEGNFSLKD
	201	NKRTLELRHG	YRYSCTPGRA	DYNQVSFQKL	NLIISTTPKL	IDPVSHRRTI
	251	STAQLIGSSN	PQHQAELMWR	ISLTVSVLLL	CLLAVPLSYF	NPRSGHTYNI
	301	LIAIGLFLIY	QNGLTLLFEA	VEDGKIHFWL	GLLPMHIIMF	VIAIVLLRVR
	351	SMPSQPFWQA	VG...			

Further work revealed the complete nucleotide sequence <SEQ ID 505>:

50	1	ATGATTTATC	AAAGAAACCT	CATCAAAGAA	CTCTCTTTTA	CCGCCGTCCG
	51	CATTTTCGTC	GTCCTCTTGG	CGGTGTTGGT	GTCCACGCAG	GCGATCAACC
	101	TGCTTGGCCG	CGCAGCTGAC	GGGCGTGTG	CCATCGATGC	CGTGTGGGCC
	151	TTAGTCGGCT	TCTGGGTCAT	CGGTATGACC	CCGCTTTTGC	TGGTGTGAC
55	201	CGCATTATC	AGCACGCTGA	CCGTATTGAC	CCGCTACTGG	CGCGACAGCG
	251	AAATGTCGGT	CTGGCTATCC	TGCGGATTGG	CGTTGAAACA	GTGGATACGC
	301	CCCGTCATGC	AGTTTGCCGT	GCCGTTTGCC	ATCCTGATTG	CCGTCATGCA
	351	GCTTTGGGTG	ATACCGTGGG	CAGAGCTGCG	CAGCCGCGAA	TATGCCGAAA
	401	TTTTGAAGCA	GAAGCAGGAA	TTGTCTTTGG	TGGAAGCCGG	CGAGTTCAAT
	451	AACTTGGGCA	AGCGCAACGG	CAGgggtttaT	Ttcgtcgaaa	CCTTTGACAC
60	501	CGaatccgGC	ATCATGAAAA	ACCTGTtcct	GcGCGAACAG	GACAAAAACG
	551	gcggcgacaA	CATCATCTTC	GCcaaaGAag	gtaactTctc	gctgaaggac

10

15

20

25

30

35

40

45

50

55

60

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is

predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 60

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 507>:

```

5      1  ..GGTGGTGGTT TTATCAATGC TTCCTGTGCC ACTTTGACGA CAGCCAAACC
      51  GCAATATCAA GCAGGAGACC TTAGCGCTTT TAAGATAAGG CAAGGCAATG
     101  TTGTAATCGC CGGACACGGT TTGGATGCAC GTGATACCGA TTACACACGT
     151  ATTCTCAGTT ATCATTCCAA AATCGATGCA CCCGTATGGG GACAAGATGT
     201  TCGTGTCTGC GCGGGACAAA ACGATGTGGC CGCAACAGGT GATGCACATT
    10  251  CGCCTATTCT CAATAATGCT GCTGCCAATA CGTCAAACAA TACAGCCAAC
     301  AACGGCACAC ATATCCCTTT ATTTGCGATT GATACAGGCA AATTAGGAGG
     351  TAT.GTATGC CAACAAAATC ACCTTGATCA GTACGGTCGA GCAAGCAGGC
     401  ATTCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 508; ORF113>:

```

15      1  ..GGGFINASCA TLTTAKPQYQ AGDLSAFKIR QGNVVIAGHG LDARDTDYTR
      51  ILSYHSKIDA PVWGQDVRVV AGQNDVAATG DAHSPILNNA AANTSNTAN
     101  NGTHIPLFAI DTGKLGGXVC QQNHLDQYGR ASRHS*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with with pspA putative secreted protein of *N.meningitidis* (accession AF030941)

20 ORF and pspA show 44% aa identity in 179aa overlap:

```

      orf113  GGGFINASCATLTAKPQYQAGDLSAFKIRQGNVVIAGHGLDARDTDYTRILSYHSKIDA 60
      GGG INA+ TLT+ P   G+L+ F + G VVI G GLD D DYTRILS ++I+A
      pspA    GGGLINAASVTLTSGVPVLNNGNLTGFDVSSGKVVIIGKGLDTSADYTRILSRAAEINA 256

25      orf113  PVWGQDVRVVAGQNDVAATGDAHSPILXXXXXXXXXXXXXXXXGTHIPLFAIDTGKLGGMYA 120
      VWG+DV+VV+G+N +   G                               + P AIDT LGGMYA
      pspA    GVWGKDVKVVS GKNKLD FDG-----SLAKTASAPSSSDSVTPTVAIDTATLGGMYA 307

30      orf113  NKITLISTVEQAGIRNQGWFFASAGNVAVNAEGKLVNTGMIAATGENHAVSLHARNVHN 179
      +KITLIST  A IRN+G+ FA+ G V ++A+GKL N+G I A       +++ A+ V N
      pspA    DKITLISTDNGAVIRNKGRIFAATGGVTLSDGKLSNSGSIDAA----EITISAQTVDN 362

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF113 shows 86.5% identity in 52aa overlap at the N- terminal part and 94.1% identity in 17aa

35 overlap at the C-terminal part with a predicted ORF (ORF113ng) from *N. gonorrhoeae*:

```

      orf113                                     GGGFINASCATLTAKPQYQAGDLSAFKIR   30
      orf113ng  SHPSQLNGYIEVGGRRAEVVIANPAGIAVNGGGFINASRATLTGQPQYQAGDFSGFKIR  224
      orf113    ||||| |||||:|||||:|:||||
      orf113ng  QGNVVIAGHGLDARDTDYTRILSYHSKIDAPVWGQDVRVVAGQNDVAATGDAHSPILNNA  90
      orf113    |||:||||| |||||:||||
      orf113ng  QGNAVIAGHGLDARDTDFTRILVCQQNHLDQYGRTSRHS  263
      orf113    IDTGKLGGXVCQQNHLDQYGRASRHS  135
      orf113ng  DFSGFKIRQNAVIAGHGLDARDTDFTRILVCQQNHLDQYGRTSRHS  263

```

The complete length ORF113ng nucleotide sequence <SEQ ID 509> is predicted to encode a protein having amino acid sequence <SEQ ID 510>:

```

50      1  MNKTLRVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIPTH
      51  SKAFCFSALG FSLCLALGTV NIAFADGIIT DKAAPKTQQA TILQTGNGIP

```

-300-

```

101 QVNIQTPTSA GVSVNQYAQF DVGNRGAILN NSRSNTQTQL GGWIQGNPWL
151 TRGEARVVVN QINSSHPSQL NGYIEVGGRR AEVVIANPAG IAVNGGGFIN
201 ASRATLTTGQ PQYQAGDFSG FKIRQGNVAVI AGHGLDARDT DFTRILVCQQ
251 NHLDQYGRTS RHS*

```

- 5 Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 61

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 511>:

```

10      1  ..TCAACGGGAC ATAGCGAACA AAATTACACT TTGCCGCGAG AAATCACACG
      51  CAACATTTCA CTGGGTTTCAT TTGCCTATGA ATCGCATCGC AAAGCATTAA
     101  GCCATCATGC GCCCAGCCAA GGCACGTAGT TGCCGCAAAG CAACGGTATT
     151  TCGCTACCCT ATACGTCCAA TTCTTTTACC CCATTACCCA GCAGCAGCTT
     201  ATACATTATC AATCCTGTCA ATAAAGGCTA TCTTGTTGAA ACCGATCCAC
     251  GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GctGGACAGC
    15   301  CTCAAACTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA
     351  CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
     401  GTTTAGAcGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT
     451  AATGGCGCGA CTGCGGCACG TTcGATGAAT CTCAGCGTTG GCATTGCATT
     501  AAGTGCCGAG CAAGTAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
    20   551  AAAAAAGAAGT TAAGCTTCCT GATGGCGGCA CACAAACCGT ATTGGTGCCA
     601  CAGGTTTATG TACGCGTTAA AAATGGCGAC ATAGACGGTA AAGGTGCATT
     651  GTTGTCAGGC AGCAATACAC AAATCAATGT TTCAGCGAGC CTGAAAAACT
     701  CAGGCACGAT TGCAGGgCGC AATGCGCTTA TTATCAATAC CGATACGCTA
     751  GACAATATCG GTGGCGGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
    25   801  ACAAGACATC AATAATATTG GCGGCATGCT TTCTGCCGAA CAGACATTAT
     851  TGCTCAACGC AGGCAACAAC ATCAACAGCC AAAGCACCAC CGCCAGCAGT
     901  CAAAATACAC AAGGCAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA
     951  TATCACAGGC AAAGAAAAAG GTGTTT..

```

This corresponds to the amino acid sequence <SEQ ID 512; ORF115>:

```

30      1  ..STGHSEQNYT LPREITRNI LGSFAYESHR KALSHHAPSQ GTELPQSNIGI
      51  SLPYTSNSFT PLPSSSLYII NPVNKGylVE TDPRFANYRQ WLGSdYMLDS
     101  LKLDpNNLHK RLGDgYyEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
     151  NGATAARSMN LSVGIALSAE QVAQLTSDIV WLvQKEVKLP DGGTQTvLVP
    35   201  QVYVRVKNGD IDGKGALLSG SNTQINVSgs LKNSGTIAGR NALIINTDL
     251  DNIGGRIHAQ KSAVTATQDI NNIGGMLSAE QTLLNAGNN INSQSTTASS
     301  QNTQGSSTYL DRMAGIYITG KEKGV..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the *pspA* putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF115 and *pspA* protein show 50% aa identity in 325aa overlap:

```

40      Orf115: 1  STGHSEQNYTLPREITRNI LGSFAYESHRKALSHHAPSQGTELPQSNIGISLPYTSNSFT 60
      STG+S  Y  E++ +I +G AY+ +  +  P  +  NGI  +T
      pspA:  778  STGYSRSPYEPAPEVS-SIRMGISAYKGYAPQASDIPGTVPVVAENGIHPTFT----- 831

      Orf115: 61  PLPSSSLYIINPVNKGylVETDPRFANYRQWLGSdYMLDSLKLDpNNLHKRLGDgYyEQR 120
      LP+SSL+ I P NKGyl+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDgYyEQ+
    45   pspA:  832  -LPNSSLFAIAPNNKGYLIETDPAFTDYRKWLGSgYMLAALQQDPNHIHKRLGDgYyEQK 890

      Orf115: 121  LINEQIAELTGHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIV 180
      L+NEQIA+LTG+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQVA+LTSDIV
    50   pspA:  891  LVNEQIAKLTYRRLDGYTNDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIV 950

      Orf115: 181  WLvQKEVKLPDGGTQTvLVPQVYVRVKNGDIDGKGALLSGSNTQINVSgsLKN-SGTIAG 239
      WL  + V LPDG TQTVL P+VYVR +  D++G+GALLSGS  I  SG+++N  G IAG
    55   pspA:  951  WLENETVTLPDGTQTvLKPVYVRARPKDMNGQALLSGSVVDIG-SGAIENRGGLIAG 1009

```

-301-

Orf115: 240 RNALIINTDTLDNIGGRIHAQKSAVTATQDINNIGGMLSAEQTLNAGXXXXXXXXXXXX 299  
 R ALI+N + N+ G + + A DI N G + AE LLL A  
 pspA: 1010 REALILNAQNIKNLQGDLOGKNIFAAAGSDITNTGS-IGAENALLLKASNNIESRSETRS 1068

5 Orf115: 300 XXXXXXXXXXXYLD RMAGIYITGKEKG 324  
 + R+AGIY+TG++ G  
 pspA: 1069 NQNEQGSVRNIGRVAGIYLTGRQNG 1093

### Homology with a predicted ORF from *N.gonorrhoeae*

10 ORF115 shows 91.9% identity over a 334aa overlap with a predicted ORF (ORF115ng) from *N.gonorrhoeae*:

	orf115.pep	STGHSEQNYTLPREITRNISLGSFAYESHK	31
		:    :	
15	orf115ng	NEQTFGEKKVFSENGKLHNYWRARRKGHDETGHRQNYTLPEEITRDISLGSFAYESHK	71
	orf115.pep	ALSHHAPSQGTLPQSN-----GISLPYTSNSFTPLPSSSLYIINPVNKGYLVE	81
	orf115ng	ALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYIINPANKGYLVE	131
20	orf115.pep	DPRFANYRQWLGS DYMLDSLKLDPNNLHKRLGDGYEQRLINEQIAELTGHRRLDGYQND	141
	orf115ng	DPRFANYRQWLGS DYMLGSLKLDPNNLHKRLGDGYEQRLINEQIAELTGHRRLDGYQND	191
25	orf115.pep	EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVWLQKEVKLPDGGTQTVLVPQ	201
	orf115ng	EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLQKEVKLPDGGTQTVLMPQ	251
	orf115.pep	VYVRVKNGDIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	261
30	orf115ng	VYVRVKNGGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	311
	orf115.pep	SAVTATQDINNIGGMLSAEQTLNAGNNINSQSTTASSQNTQGSSTYLDRMAGIYITGK	321
35	orf115ng	SAVTATQDINNIGGILSAEQTLNAGNNINNQSTAKSSQNAQGSSTYLDRMAGIYITGK	371
	orf115.pep	EKGV	325
	orf115ng	EKGVLAAQAGKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTKGYQEIHFADNHTIR	431

An ORF115ng nucleotide sequence <SEQ ID 513> was predicted to encode a protein having amino acid sequence <SEQ ID 514>:

	1	MLVQTEKDGL	HNEQTFGEKK	VFSENGKLHN	YWRARRKGHD	ETGHREQNYT
	51	LPEEITRDIS	LGSFAYESH	KALSRHAPSQ	GTELPQSNRD	NIRTAKSNGI
	101	SLPYTPNSFT	PLPGSSLYII	NPANKGYLVE	TDPRFANYRQ	WLGS DYMLGS
45	151	LKLDPNNLHK	RLGDGYEQR	LINEQIAELT	GHRRLDGYQN	DEEQFKALMD
	201	NGATAARSMN	LSVGIALSAE	QAAQLTSDIV	WLQKEVKLP	DGGTQTVLMP
	251	QVYVRVKNGG	IDGKGALLSG	SNTQINVSGS	LKNSGTIAGR	NALIINTDTL
	301	DNIGGRIHAQ	KSAVTATQDI	NNIGGILSAE	QTLNAGNN	INNQSTAKSS
	351	QNAQGSSTYL	DRMAGIYITG	KEKGVLAQA	GKDINIIAGQ	ISNQSDQGQT
	401	RLQAGRDINL	DTVQTKGYQE	IHFADNHTI	RGSTNEVGSS	IQTGKDVTLL
50	451	SGNNLNAKAA	EVGSAKGTLA	VYAKNDITIS	SGIHAGQVDD	ASKHTGRSGG
	501	GNKLVITDKA	QSHHETAQSS	TFEGKQVVLQ	AGNDANILGS	NVISDNGTRI
	551	QAGNHVRIGT	TQTQSQSEY	HQTQKSLMS	AGIGFTIGSK	TNTQENQSQS
	601	NEHTGSTVGS	LKGDITIVAS	KHYEQTGSNV	SSPEGNNLIS	TQSM DIGAAQ
55	651	NQLNSKTTQT	YEQKGLTVAF	SSPVDLAQQ	AI AVAHKAAK	QFDKAKTTAL
	701	MPWRLPMQVG	RLFKQAKAPK	K*		

Further work revealed the following partial gonococcal DNA sequence <SEQ ID 515>:

	1	TTGCTTGTGC	AAACAGAAAA	AGACGGTTTG	CATAACGAGC	AAACCTTTGG
	51	CGAGAAGAAA	GTCTTCAGCG	AAAATGGTAA	GTTGCACAAC	TACTGGCGTG
	101	CCGCTCGTAA	AGGACATGAT	GAAACAGGGC	ATCGTGAACA	AAATTATACT
60	151	TTGCCGGAGG	AAATCACACG	CGACATTTC	CTGGGTTCAT	TGCGCTATGA
	201	ATCGCATAGC	AAAGCATTAA	GCCGTCATGC	GCCAGCCAA	GGCACTGAGT
	251	TGCCACAAAG	TAACCGGGAT	AATATCCGTA	CTGCGAAAAG	CAACGGTATT

-302-

301 TCGCTACCTT ATACGCCCAA TTCTTTTACC CCATTACCCG GCAGCAGCTT  
 351 ATACATTATC AATCCTGCCA ATAAAGGCTA TCTTGTGAA ACCGATCCAC  
 401 GCTTTGCCAA CTACCGTCAA TGGTGGGTA GTGACTATAT GCTGGGCAGC  
 451 CTCAAACCTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA  
 501 CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGCATCGTC  
 551 GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT  
 601 AATGGCGCGA CTGCGGCACG TTCGATGAAT CTCAGCGTTG GCATTGCATT  
 651 AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC  
 701 AAAAAGAAGT TAAACTTCCT GATGGCGGCA CACAAACCGT ATTGATGCCA  
 751 CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT  
 801 GTTGTCTAGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT  
 851 CAGGCACGAT TGCAGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA  
 901 GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC  
 951 ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT  
 1001 TGCTCAATGC GGGTAACAAC ATCAACAACC AAAGCACGGC CAAGACGAGT  
 1051 CAAATGCGAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA  
 1101 TATCACAGGC AAAGAAAAAG GTGTTTtagc AGCGCAGGCA GGCAAAGACA  
 1151 TCAACATCAT TGCCGTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC  
 1201 CGGCTGCAGG CAGGACGCGA CATTAACTG GATACGGTAC AAACCGGCAA  
 1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAA  
 1301 CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCCttatTG  
 1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAGG  
 1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC  
 1451 ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC  
 1501 GGTAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC  
 1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG  
 1601 ATGCCAACAT CTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT  
 1651 CAAGCAGGCA ATCATGTTTCG CATTGGTACA ACCCAAACCTC AAAGCCAAAG  
 1701 CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG  
 1751 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAACCA ATCCCAAGC  
 1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT  
 1851 TGTTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG  
 1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA  
 1951 AACCATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC  
 2001 GGTGGCATT CAGTTGCCCC TTACCGATT GGCACAACAA GCGATTGCCG  
 2051 TAGCACACAA AGCAGCAAAAC AAGTCGGACA AAGCAAAAC GACCGCGTTA  
 2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA  
 2151 GCGCACAAA ACTTAG

This corresponds to the amino acid sequence <SEQ ID 516; ORF115ng-1>:

40 1 LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT  
 51 LP EEITRDIS LGSFAYESH KALSRHAPSQ GTELPQSNRD NIRTAKSNGI  
 101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPFRFANYRQ WLGSDYMLGS  
 151 LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD  
 201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVQKEVKLP DGGTQTVLMP  
 45 251 QVYVRVKNNG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL  
 301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNLLNAGNN INNQSTAKSS  
 351 QNAQGSSTYL DRMAGIYITG KEGVLAAQA GKDINIIAGQ ISNQSDQGGT  
 401 RLQAGRDLNL DTVQTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTLT  
 451 SGNNLNAAKAA EVGSAKGTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG  
 50 501 GNKLVIDDKA QSHHETAQSS TFEKGQVVLQ AGNDANILGS NVISDNTRI  
 551 QAGNHVRIGT TQTQSQSEY HQTQKSLMS AGIGFTIGSK TNTQENQSQS  
 601 NEHTGSTVGS LKGDPTIVAS KHYEQTGSNV SPEGNNLIS TQSM DIGAAQ  
 651 NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAAN KSDKAKTTAL  
 701 MPWRLPMQVG RPIKQAKAHK T\*

55 This gonococcal protein (ORF115ng-1) shows 91.9% identity with ORF115 over 334aa:

20 30 40 50 60 70  
 orf115ng-1.p NEQTFGEKKVFSENGKLHN YWRARRKGHD ETGHREQNYT LP EEITRDIS LGSFAYESH SK  
 orf115 STGHSEQNYTLP REITRNIS LGSFAYESH RK  
 10 20 30  
 80 90 100 110 120 130  
 orf115ng-1.p ALSRHAPSQGT ELPQSNRD NIRTAKSNGI SLPYTPNSFT PLPGSSLYII NPANKGYLVE T  
 orf115 ALSHHAPSQGT ELPQSN-----GISLPYTSNSFT PLPSSSLYII NPVKNKYLVET  
 40 50 60 70 80

-303-

```

      140      150      160      170      180      190
orf115ng-1.p DPRFANYRQWLGS DYMLGSLKLD PNNLHKRLG DGYEQRLNEQIAELTGHRRLDGYQND
      |||
orf115      DPRFANYRQWLGS DYMLDSLKLD PNNLHKRLG DGYEQRLNEQIAELTGHRRLDGYQND
      90      100      110      120      130      140

      200      210      220      230      240      250
orf115ng-1.p EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWL VQKEVKLPDGGTQT VLVMPQ
      |||
orf115      EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVWL VQKEVKLPDGGTQT VLVMPQ
      150      160      170      180      190      200

      260      270      280      290      300      310
orf115ng-1.p VYVRVKNGGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK
      |||
orf115      VYVRVKNGDIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK
      210      220      230      240      250      260

      320      330      340      350      360      370
orf115ng-1.p SAVTATQDINNIGGILSAEQ TLLNAGNNINNQSTAKSSQNAQGSSTYLDRMAGIYITGK
      |||
orf115      SAVTATQDINNIGGMLSAEQ TLLNAGNNINSQSTASSQNTQGSSTYLDRMAGIYITGK
      270      280      290      300      310      320

      380      390      400      410      420      430
orf115ng-1.p EKGVLAAQAGKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQEIHFDADNHTIR
      |||
orf115      EKG

```

In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

```

30 gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length
   = 2273
   Score = 604 bits (1541), Expect = e-172
   Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

35 Query: 1 LLVQTEKDGLHNEQTFGEKKVFSENGKLHNYWRARRKGHDETG HREQNYTLPEEITRDIS 60
      L+V T + L N++T G K + ++ G LH Y R +KG D TG+ Y E++ I
Sbjct: 739 LIVGTPESALDNDETLGTKTI-TDKGDLHRYHRHKKGRDSTGYRSRSPYEPAPEVS-SIR 796

40 Query: 61 LGSFAYESHKALSRHAPSQGTTELPOSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120
      +G AY+ + AP Q +++P + + NGI +T LP SSL+ I
Sbjct: 797 MGISAYKGY-----APQQASDIPGTV---VPVVAENGIHPTFT-----LPNSSLFAI 840

45 Query: 121 NPANKGYLVETDPRFANYRQWLGS DYMLGSLKLD PNNLHKRLG DGYEQRLNEQIAELT 180
      P NKGYL+ETDP F +YR+WLGS YML +L+ DEN++HKRLG DGYEQ+L+NEQIA+LT
Sbjct: 841 APNNKGYLIETDPAFTDYRKWLGS GYMLAALQQDPNHIHKRLG DGYEQKL VNEQIAKLT 900

50 Query: 181 GHRRLDGYQND EEFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWL VQKEVKLP 240
      G+RRLDGY NDEEFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP
Sbjct: 901 GYRRLDGYTNDEEFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIVWLENETVTL P 960

55 Query: 241 DGGTQT VLVMPQVYVRVKNGGIDGKGALLSGSNTQINVSGSLKN-SGTIAGRNALIINTDT 299
      DG TQT V L P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N
Sbjct: 961 DGT TQT V L K P K VYVRARPKDMNGQ GALLSGSVVDIG-SGA IENRGGLIAGREALILNAQN 1019

60 Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQ TLLNAGNNINNQSTAKSSQNAQGSSTY 359
      + N+ G + + A DI N G I AE LLL A NNI ++S +S+QN QGS
Sbjct: 1020 IKNLQGD LQGNIFAAAGSDITNTGSI-GAENALLKASNIESRSETRSNQNEQGSVRN 1078

65 Query: 360 LDRMAGIYITGKEKGVLAQAGKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQ 419
      + R+AGIY+TG++ G + AG +I + A +++NQS+ GQT L AG DI DT + Q
Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSDTTGISRNQ 1138

70 Query: 420 EIHFDADNHTIRGSTNEVGSSSIQTKGDVTL TSGNNLNAKAAEVGSAKGT LAVYAKNDITI 479
      FD+DN+ IR NEVGS+I+T+G+++L + ++ +AAEVGS +G L + A DI +
Sbjct: 1139 NTIFDS DNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQGRKLAAAGRDIKV 1198

Query: 480 SSGIHAGQVDDASKHTGRSGGGNKLIVITDKAQSHHETAQSSTFEGKQVV LQAGNDANILG 539
      +G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G
Sbjct: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQNGQAVSGTLDGKEIILVSGRDI TVTG 1258

```

Query: 540 SNVISDNGTRIQAQNHVRIGTTQTQSQSEYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598  
 SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S  
 Sbjct: 1259 SNIIADNHTILSAKNNIVLKAETRSTSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNRNRS 1318

5 Query: 599 QSNEHTGSTVGSGLKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSMIDGAAQNQLNSKTT 658  
 ++ HT S VGS L G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++  
 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISSGKISIDAAQNRYSQESK 1378

10 Query: 659 QTYEQKGLTVAFSSPVT 676  
 Q YEQKG+TVA S PV +  
 Sbjct: 1379 QVYEQKGVTVAISVPVVN 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 62

15 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 517>:

1 ..TCAGGGAATA ACCTCAATGC CAAAGCTGCC GAAGTCAGCA GCGCAAACGG  
 51 TACACTCGCT GTGTCTGCCA ATAATGACAT CAACATCAGC GCAGGCATCA  
 101 ACACGACCCA TGTGTATGAT GCGTCCAAAC ACACAGGCAG AAGCGGTGGT  
 151 GGCAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACCGC  
 201 CCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG  
 251 ATGCCAACAT CCTTGGCAGC AATGTTATTT CCGATAATGG CACCCAGATT  
 301 CAAGCAGGCA ATCATGTTTCG CATTGGTACA ACCCAAATC AAAGCCAAAG  
 351 CGAAACCTAT CATCAAACCC AGAAATCAGG ATTGATGAGT GCAGGTATCG  
 401 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC  
 25 451 AACGAACATA CAGGCAGTAC CGTAGGCAGC TTGAAAGGCG ATACCACCAT  
 501 TGTTGCAGGC AAACACTACG AACAAATCGG CAGTACCGTT TCAGCCCCGG  
 551 AAGGCAACAA TACCATCTAT GCCCAAAGCA TAGACATTCA AGCGGCACAC  
 601 AACAAATTAA ACAGTAATAC CACCCAAACC TATGAACAAA AAGG.CTAAAC  
 651 GGTGGCATT CAGTTGCCCG TTACCGATT GGCACAACAA ...

30 This corresponds to the amino acid sequence <SEQ ID 518; ORF117>:

1 ..SGNNLNAAKAA EVSSANGTLA VSANNDINIS AGINTTHVDD ASKHTGRSGG  
 51 GNKLVIDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTQI  
 101 QAGNHVRIGT TQTQSQSEY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS  
 151 NEHTGSTVGS LKGDTTIVAG KHYEQIGSTV SSPEGNNTIY AQSIDIQAAH  
 35 201 NKLNSNTTQT YEQKXLTVAF SSPVTDLAQQ ...

Computer analysis of this amino acid sequence gave the following results:

Homology with the *pspA* putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF117 and *pspA* protein show 45% aa identity in 224aa overlap:

Orf117: 4 NLNAKAAEVSSANGTLAVSANNDINISAGINTTHVDDASKHTGRSGGGNKLVIDKAQSH 63  
 ++ +AAEV S G L ++A DI + AG T +DA K+TGRSGGG K +T ++  
 pspA: 1173 DIRIRAAEVGSEQRLKLAAGRDIKVEAGKAHTETEDALKYTGRSGGGIKQKMTRHLKNQ 1232

Orf117: 64 HETAQSSTFEGKQVVLQAGNDANILGSNVISDNGTQIQAGNHVRIGTTQTQSQSEYHQT 123  
 + A S T +GK+++L +G D + GSN+I+DN T + A N++ + +T+S+S ++  
 45 pspA: 1233 NGQAVSGTLDGKEIILVSGRDITVTGSNIIADNHTILSAKNNIVLKAETRSTSAEMNKK 1292

Orf117: 124 QKSGLM-SAGIGFTIGSKTNTQENQSQSEHTGSTVGSGLKGDTTIVAGKHYEQIGSTVSS 182  
 +KSGLM S GIGFT GSK +TQ N+S++ HT S VGS L G+T I AGKHY Q GST+SS  
 50 pspA: 1293 EKSGLMGSGGIGFTAGSKKDTQTNRSETVSHTESVVGSLNGNTLISAGKHYTQTGSTISS 1352

Orf117: 183 PEGNNTIYAQSIDIQAAHNKLNSNTTQTYEQKXLTVAFSSPVT 226  
 P+G+ I + I I AA N+ + + Q YEQK +TVA S PV +  
 pspA: 1353 PQGDVGISSGKISIDAAQNRYSQESKQVYEQKGVTVAISVPVVN 1396



Homology with a predicted ORF from *N.gonorrhoeae*

ORF117 shows 90% identity over a 230aa overlap with a predicted ORF (ORF117ng) from *N.gonorrhoeae*:

5	orf117.pep	SGNNLNAKAAEVSSANGTLAVSANNDINIS	30
	orf117ng	IHFDADNHTIRGSTNEVGSSIQTKGDTVLLSGNNLNAKAAEVGSAKGT LAVYAKNDITIS	480
	orf117.pep	AGINTTHVDDASKHTGRSGGKNLVIDKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	90
10	orf117ng	SGIHAGQVDDASKHTGRSGGKNLVIDKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	540
	orf117.pep	NVISDNGTQIQAGNHVRIGTTQTQSQSEYHQTQKSGLMSAGIGFTIGSKTNTQENQSQS	150
15	orf117ng	NVISDNGTRI QAGNHVRIGTTQTQSQSEYHQTQKSGLMSAGIGFTIGSKTNTQENQSQS	600
	orf117.pep	NEHTGSTVGS LKGD TTIVAGKH YEIQIGSTVSSPEGNN TIYAQSIDIQA AHNKLN SNTTQT	210
	orf117ng	NEHTGSTVGS LKGD TTIVASKH YEQTGS NVSSPEGNN LISTQSM DIGAAQNQLNSKTTQT	660
20	orf117.pep	YEQKXLTVAFSSPVTDLAQQ	230
	orf117ng	YEQKGLTVAFSSPVTDLAQQAIAVAHKA AKQFDKAKTTALMPWR LPMQVGR LFKQAKAPK	720

An ORF117ng nucleotide sequence <SEQ ID 519> was predicted to encode a protein having amino acid sequence <SEQ ID 520>:

25	1	..LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT
	51	LPEEITRDIS LGSFAYESH S KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
	101	SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGS DYMLGS
	151	LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
	201	NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVOKEVKLP DGGTQTVLMP
30	251	QVYVRVKNGG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
	301	DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNAGN INNQSTAKSS
	351	QNAQGSSTYL DRMAGIYITG KEGVLAAQA GKDINIIAGQ ISNQSDQGQT
	401	RLQAGRDINL DTVQTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDTVLL
	451	SGNNLNAKAA EVGSAKGT LA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
35	501	GNKLVIDKA QSHHETAQSS TFEKQVVLQ AGNDANILGS NVISDNGTRI
	551	QAGNHVRIGT TQTQSQSEY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS
	601	NEHTGSTVGS LKGD TTIVAS KH YEQTGS NV SSPEGNN LIS TQSM DIGAAQ
	651	NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKA AK QFDKAKTTAL
	701	MPWRLPMQVG RLFQAKAPK K*

40 Further work revealed the following gonococcal partial DNA sequence <SEQ ID 521>:

	1	TTGCTTGTGC AAACAGAAAA AGACGGTTTG CATAACGAGC AAACCTTTGG
	51	CGAGAAGAAA GTCTTCAGCG AAAATGGTAA GTTGCAACAAC TACTGGCGTG
	101	CGCGTCGTAA AGGACATGAT GAAACAGGGC ATCGTGAACA AAATTATACT
	151	TTGCCGGAGG AAATCACACG CGACATTTCA CTGGGTTTCA TTGCCTATGA
45	201	ATCGCATAGC AAAGCATTA GCGTCATGC GCCCAGCCAA GGCACTGAGT
	251	TGCCACAAAG TAACCGGGAT AATATCCGTA CTGCGAAAAG CAACGGTATT
	301	TCGCTACCCT ATACGCCCAA TTCTTTTACC CCATTACCCG GCAGCAGCTT
	351	ATACATTATC AATCTGCCA ATAAAGGCTA TCTTGTGTA ACCGATCCAC
	401	GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGGCAGC
50	451	CTCAAACTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA
	501	CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
	551	GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTAAAGC CTTAATGGAT
	601	AATGGCGCGA CTGCGGCACG TTCGATGAAT CTCAGCGTTG GCATTGCATT
	651	AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
55	701	AAAAAGAAGT TAAACTTCCT GATGGCGGCA CACAAACCGT ATTGATGCCA
	751	CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT
	801	GTTGTCAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
	851	CAGGCACGAT TGCAGGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA
	901	GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
60	951	ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT
	1001	TGCTCAATGC GGGTAACAAC ATCAACAACC AAAGCACGGC CAAGAGCAGT
	1051	CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA

5  
 10  
 15  
 20

```

1101 TATCACAGGC AAAGAAAAAG GTGTTTTAGC AGCGCAGGCA GGCAAAGACA
1151 TCAACATCAT TGCCGGTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC
1201 CGGCTGCAGG CAGGACGCGA CATTAACTG GATACGGTAC AAACCGGCAA
1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAA
1301 CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCCtatTG
1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAAGG
1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC
1451 ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC
1501 GGTAAATAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC
1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
1601 ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT
1651 CAAGCAGGCA ATCATGTTTCG CATTGGTACA ACCCAAATC AAAGCCAAAG
1701 CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG
1751 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC
1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT
1851 TGTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG
1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA
1951 AACCAATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC
2001 GGTGGCATTG AGTTCGCCCG TTACCGATTG GGCACAACAA GCGATTGCCG
2051 TAGCACACAA AGCAGCAAAAC AAGTCGGACA AAGCAAAAAC GACCGCGTTA
2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA
2151 GGCGCACAAA ACTTAG
  
```

This corresponds to the amino acid sequence <SEQ ID 522; ORF117ng-1>:

25  
 30  
 35

```

1 LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT
51 LP EEITRDIS LGSFAYESH KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGS DYMLGS
151 LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLQKEVKLP DGGTQTVLMP
251 QYVVRVKNNG IDGKGALLSG SNTQINVS GS LKNSGTIAGR NALIINTDTL
301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTL LNAGNN INNQSTAKSS
351 QNAQGSSTYL DRMAGIYITG KEKGVLAQA GKDINIAGQ ISNQSDQGQT
401 RLQAGRDINL DTVQTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTLL
451 SGNNLNAKAA EVGSAKGT LA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
501 GNKLVIIDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTRI
551 QAGNHVRIGT TQTQSQSEY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS
601 NEHTGSTVGS LKGDTTIVAS KHYEQTGSNV SSPEGNLIS TQSM DGAAG
651 NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAAN KSDKAKTTAL
701 MPWRLPMQVG RPIKQAKAHK T*
  
```

ORF117ng-1 shows the same 90% identity over a 230aa overlap with ORF117. In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

45  
 50  
 55  
 60  
 65

```

gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis]Length =
2273
Score = 604 bits (1541), Expect = e-172
Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

Query: 1 LLVQTEKDGLHNEQTFGEKKVFSENGKLHNYWRARRKGHD ETGHREQNYTLPEEITRDIS 60
L+V T + L N++T G K + ++ G LH Y R +KG D TG+ Y E++ I
Sbjct: 739 LIVGTPESALDND ETLGTKTI-TDKGDLHRYHRHHKKGRDSTGYSRSPYEPAPEVS-SIR 796

Query: 61 LGSFAYESHKALSRHAPSQGT ELPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120
+G AY+ + AP Q +++P + + NGI +T LP SSL+ I
Sbjct: 797 MGISAYKGY-----APQQASDIPGTV---VPVVAENGIHPTFT-----LPNSSLFAI 840

Query: 121 NPANKGYLVETDPRFANYRQWLGS DYMLGSLKLDPNNLHKRLGDGYEQR LINEQIAELT 180
P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQR+L+NEQIA+LT
Sbjct: 841 APNNKGYLIETDPAFTDYRKWLGS GYMLAALQQDPNHIHKRLGDGYEQRKLVNEQIAKLT 900

Query: 181 GHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLQKEVKLP 240
G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP
Sbjct: 901 GYRRLDGYTNDEEQFKALMDNGITIAKELQLT PGIALSAEQVARLTSDIVWLENETVTL P 960

Query: 241 DGGTQTVLMPQYVVRVKNNGIDGKGALLSGSNTQINVS GSKN-SGTIAGR NALIINTDT 299
DG TQTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N
Sbjct: 961 DGT TQTVLKP KYVVRARPKDMNGQALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019

Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTL LNAGNNINNQSTAKSSQNAQGSSTY 359
  
```

-307-

+ N+ G + + A D I N G I AE LLL A NNI ++S +S+QN QGS  
 Sbjct: 1020 IKNLQGDLOQKNIFAAAGSDITNTGSI-GAENALLLKASNNIESRSETRSNQNEQGSVRN 1078  
 Query: 360 LDRMAGIYITGKEKGVLAQAQAGKDINIIAGQISNQSDQGGQTRLQAGRNDINLDTVQTGKYQ 419  
 + R+AGIY+TG++ G + AG +I + A +++NQs+ GQT L AG DI DT + Q  
 Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEdGQTVLNAGGDIRSDTTGISRNQ 1138  
 Query: 420 EIHFDADNHTIRGSTNEVGSSIQTKGDVTLISGNNLNAAAEVGSAGKTLAVYAKNDITI 479  
 FD+DN+ IR NEVGS+I+T+G+++L + ++ +AAEVGS +G L + A DI +  
 Sbjct: 1139 NTIFDSNDYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQGRLKLAAGRDIKV 1198  
 Query: 480 SSGIHAGQVDDASKHTGRSGGKNLVIDKAQSHHETAQSSTFEGKQVVLQAGNDANILG 539  
 +G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G  
 Sbjct: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQNGQAVSGTLDGKEIILVSGRDITVTG 1258  
 Query: 540 SNVISDNTRIQAQGNHVRIGTTQTQSQSETYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598  
 SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S  
 Sbjct: 1259 SNIIADNHTILSAKNNIVLKAETRSRAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNRs 1318  
 Query: 599 QSNEHTGSTVGLKGDTTIVASKHYEQTSNVSSPEGNNLISTQSMIDIGAAQNQLNSKTT 658  
 ++ HT S VGSL G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++  
 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISSGKISIDAAQNRYsQESK 1378  
 Query: 659 QTYEQKGLTVAFSSPVTD 676  
 Q YEQKG+TVA S PV +  
 Sbjct: 1379 QVYEQKGVTVAISVPVVN 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 63

30 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 523>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTGCGCTA  
 51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCG  
 101 GAACTCCGA CAAAGATGCC CTGCTCAACA GCAWAACCAG CCATGTCCGC  
 151 GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCCAACCggc  
 201 GGTCAAAAAA ACGGCAAAAC CCAAGACCC CGyCATGCGC AACCTGCAAG  
 251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG  
 301 TTCAAACCG AAATCGAAAC CGCCTTGGA GAAAGCGGCA TTATCGGCAA  
 351 CTCCGCCAC ACCGTTTCCG AACCCTAAAC CGGACATTCC GCAACGAAAC  
 401 CTGCCGACGC GTCGGCAAAA CCTGCACCCG TTCCGCAAAAC ACCTGCAAAA  
 451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAT CCTGGTTTGA  
 501 CGTGCGCATC GACTTCATCT CCTAT...

This corresponds to the amino acid sequence <SEQ ID 524; ORF119>:

1 MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSXTSHVR  
 51 DGKPSGGSVMPKPPAVKK TAKPQDPXMR NLQEQDAVYI AKQKQAKASP  
 101 FKTEIETALE ESGIIGNSAH TVSEPQTGHS ATPADASAK PAPVPQTPAK  
 151 PLITLKELSK VELSWFDVRI DFISY...

Further work revealed the complete nucleotide sequence <SEQ ID 525>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTGCGCTA  
 51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCG  
 101 GAACTCCGA CAAAGATGCC CTGCTCAACA GCAAAACCAG CCATGTCCGC  
 151 GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCCAACCggc  
 201 GGTCAAAAAA ACGGCAAAAC CCAAGACCC CGCATGCGC AACCTGCAAG  
 251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG  
 301 TTCAAACCG AAATCGAAAC CGCCTTGGA GAAAGCGGCA TTATCGGCAA  
 351 CTCCGCCAC ACCGTTTCCG AACCCTAAAC CGGACATTCC GCACCGAAAC  
 401 CTGCCGACGC GCGGGCAAAA CCTGCACCCG TTCCGCAAAAC ACCTGCAAAA  
 451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAC CCTGGTTTGA  
 501 CGTGCGCTTC GACTTCATCT CCTATATCGC GCTGACCGAA GCCAAGAAGC  
 551 TGCACGCACT GCCGCGCCTT TCCAACCGCT GCCGCTACCA GATTGTCGGC  
 601 TGCACCATGG ACGACCATTT CCAGATTGCC GAACCATCC CGGGCATCCG

-308-

5 651 CTATCAGGCA TTTATCGTGG GTATTCAGGC AGTCAGCCGC AACGGACTTG  
 701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA CGCATTTCGCA  
 751 CAAAGCATGG GCGGTCAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA  
 801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACCATCG  
 851 CCATCCATTT GGTTCCTCCG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC  
 901 GCCGTAACGG GCGTGGGTTT CGTTTTGGAA GACGACGGCG CGTTCCACTA  
 951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG  
 1001 AGCCGTTTAC CAACGCCCTT TTGGACAACC AGTCCTACAA AGGCTTCAGT  
 1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GGCGAAAAAA CCTTCGACGA  
 1101 TTTGTTTATG GATTGGCGG TACGCTGTC CGGCCAGTTG AACCTGAATC  
 1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTG  
 1201 CGCACTTATG TATTGGCGCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA  
 1251 ACCGGGCGGC AAAACCGCAT TCGCCTGTT CTCCTAA

This corresponds to the amino acid sequence <SEQ ID 526; ORF119-1>:

15 1 MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR  
 51 DGKPSGGGSM MPKPQPAVKK TAKPQDPAMR NLQEQDAVYI AKQKQAKASP  
 101 FKTEIETALE ESGIIGNSAH TVSEPQTGHS APKPADAPAK PAPVPQTPAK  
 151 PLITLKELSK VELPWFDVRF DFISYIALTE AKELHALPRL SNRCRYQIVG  
 201 CTMDDHFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA  
 20 251 QSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELLS  
 301 AVTGVGVFLE DDGAFHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFS  
 351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV  
 401 RTYVLARQSE MLKVGIEPGG KTALRLFS\*

Computer analysis of this amino acid sequence gave the following results:

#### 25 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF119 shows 93.7% identity over a 175aa overlap with an ORF (ORF119a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
30	orf119.pep	MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRD					DGKPSGGGSM
	orf119a	MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRD					DGKPSGGGPM
		10	20	30	40	50	60
35	orf119.pep	MPKPQPAVKKTA	MPKPQPAVKKTA	MPKPQPAVKKTA	MPKPQPAVKKTA	MPKPQPAVKKTA	MPKPQPAVKKTA
	orf119a	MPKPQPAVKKTA	MPKPQPAVKKTA	MPKPQPAVKKTA	MPKPQPAVKKTA	MPKPQPAVKKTA	MPKPQPAVKKTA
		70	80	90	100	110	120
40	orf119.pep	TVSEPQTGHSATKPADASAKPAPVPQTPAKPLITLKELSKVELSWFDVRF					DFISY
	orf119a	TVSEPQTGHSAPKPADAPAKPVVPQTPAKPLITLKELSKVELPWFDVRF					DFISYIALTE
		130	140	150	160	170	180
45	orf119a	AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS					
		190	200	210	220	230	240

The complete length ORF119a nucleotide sequence <SEQ ID 527> is:

50 1 ATGATTTACA TCGTACTGTT CCTCGCCGCC GTCCTCGCCG TTGTGCGCTA  
 51 CAATATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCG  
 101 GGCACCTCCGA CAAAGATGCC CTGCTCAACA GCAAACCAG CCATGTCCGC  
 151 GACGGCAAAC CGTCCGGCGG GCCAGTCATG ATGCCGAAAC CCAACCGGC  
 201 GGTCAAAAAA ACGGCAAAAT CCCAAGACCC CGCCATGCGC AACCTGCAAG  
 251 AGCAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG  
 55 301 TTCAAACCG AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGCAA  
 351 CTCCGCCAC ACCGTTCCCG AACCCCAAAC CGGACATTCC GCACCAAAAC  
 401 CTGCCGACGC GCCGGCAAAA CCTGTTCCCG TTCGCAAAAC GCCGGCAAAA  
 451 CCGCTGATTA CGCTCAAAGA GCTGTCGAAG GTCGAGCTGC CCTGGTTTGA  
 501 CGTGCGCTTC GACTTCATCT CTTATATCGC GCTGACCGAA GCCAAAGAAC  
 60 551 TGCACGCACT GCCGCGCCTT TCCAACCGCT GCCGCTACCA GATTGTCGGC  
 601 TGCACCATGG ACGACCATT TCCAGATTGCC GAACCCATCC CGGGCATCCG

-309-

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

```

651 CTATCAGGCA TTTATCGTGG GTATTCAAGC AGTCAGCCGC AACGGACTTG
701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA TGCATTGCGA
751 CACAGCATGG GCGGTCAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA
801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACTATCG
851 CCATCCATTT GGTTCCTCCG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
901 GCCGTAAACG GCGTGGGTTT CGTTTGGGAA GACGACGGCG CGTTCCACTA
951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
1001 AGCCGTTTAC CAATGCCCTT TTGGACAACC AGTCCTATAA AGGCTTCAGT
1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GCGGAAAAAA CCTTCGACGA
1101 TTTGTTTATG GATTTGGCGG TACGCTGTG CCGCCAGTTG AACCTGAATC
1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTG
1201 CGCACTTATG TATTGGCTCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
1251 ACCGGGCGGC AAAACCGCAT TGCCTGTG CTCCTAA

```

This encodes a protein having amino acid sequence <SEQ ID 528>:

15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

```

1 MIYIVLFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
51 DGKPSGGPVM MPKPQPAVKK TAKSQDPAMR NLQEQDAVYI AKQKQAKASP
101 FKTEIETALE ESGIIGNSAH TVPEPQTGHS APKPADAPAK PVPVPQTPAK
151 PLITLKELSK VELPWFDFVRF DFISYIALTE AKELHALPRL SNRCRYQIVG
201 CTMDDHFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA
251 HSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS
301 AVTGVGVFLE DDGAFHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFS
351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
401 RTYVRLARQSE MLKVGIEPGG KTALELFS*

```

ORF119a and ORF119-1 show 98.6% identity in 428 aa overlap:

25  
30  
35  
40  
45  
50  
55  
60  
65

```

orff119a.pep      10      20      30      40      50      60
MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHV RDGKPSGGPVM
|||||:|||||
orff119-1          10      20      30      40      50      60
MIYIVLFLAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHV RDGKPSGG SVM

orff119a.pep      70      80      90      100     110     120
MPKPQPAVKKTAKSQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH
|||||:|||||
orff119-1          70      80      90      100     110     120
MPKPQPAVKKTAKPQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH

orff119a.pep      130     140     150     160     170     180
TVPEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDFVRFDFISYIALTE
|| |||||:|||||
orff119-1          130     140     150     160     170     180
TVSEPTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDFVRFDFISYIALTE

orff119a.pep      190     200     210     220     230     240
AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS
|||||:|||||
orff119-1          190     200     210     220     230     240
AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS

orff119a.pep      250     260     270     280     290     300
AFNRQVDAFAHSMGGQTLHTDLAAFIIEVASALDAFCARVDQTIAIHLVSP TSISGVELRS
|||||:|||||
orff119-1          250     260     270     280     290     300
AFNRQVDAFAQSMGGQTLHTDLAAFIIEVASALDAFCARVDQTIAIHLVSP TSISGVELRS

orff119a.pep      310     320     330     340     350     360
AVTGVGVFLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA
|||||:|||||
orff119-1          310     320     330     340     350     360
AVTGVGVFLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA

orff119a.pep      370     380     390     400     410     420
GEKTFDDLFM DLAVRLSGQLNLNLVNDKMEEVSTQWLKDV RTYVRLARQSEMLKVGIEPGG
|||||:|||||
orff119-1          370     380     390     400     410     420
GEKTFDDLFM DLAVRLSGQLNLNLVNDKMEEVSTQWLKDV RTYVRLARQSEMLKVGIEPGG

```

```

orf119a.pep    KTALRLFSX
               |||||
orf119-1       KTALRLFSX

```

### 5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF119 shows 93.1% identity over a 175aa overlap with a predicted ORF (ORF119ng) from *N.gonorrhoeae*:

```

10 orf119.pep    MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSXTSHVRD GKPSGGSV 60
    |||||:|||||
orf119ng        MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRD GKPSGGPVM 60

orf119.pep      MPKPQPAVKKTAKPQDPXMRNLQEQDAVYIAKQKQAKASPFKTEIETALESIGIIGNSAH 120
    |||||:|||||
orf119ng        MPKPQPAVKKPAKQDSAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEEIGIIGNSAH 120

15 orf119.pep      TVSEPQTGHSATKPADASAKPAPVPQTPAKPLITLKELSKVELSWFDFRIDFISY 175
    |||||:|||||
orf119ng        TVSEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWDFVRFDFISYIALTE 180

```

The complete length ORF119ng nucleotide sequence <SEQ ID 529> is:

```

20 1 ATGATTTACA TCGTACTGTT CCTCGCCGCC GTCCTCGCCG TTGTCGCCTA
    51 CAATATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCG
    101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAAACCAG CCATGTCCGC
    151 GACGGCAAAC CGTCCGGCGG GCCAGTCATG ATGCCGAAAC CCCAACCGGC
    201 GGTCAAAAAA CCGGCCAAAC CCCAAGACTC CGCCATGCGC AACCTGCAAG
25 251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAA AGCCTCCCCG
    301 TTCAAAACCG AAATCGAAAC CGCCTTGGAA GAAATCGGCA TTATCGGCAA
    351 CTCCGCCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCACCGAAAC
    401 CTGCCGACGC GCGGGCAAAC CCCGTTCCCG TTCCGCAAAC GCGGGCAAAC
    451 CCGCTGATTA CGCTCAAAGA GCTGTGGAAG GTCGAGCTGC CCTGGTTTGA
30 501 CGTGCCTtc gACTTCATCT CCTATATCGC GCTGACCGAA GCCAAAGAAC
    551 TGCACGCACT GCGCGCCTT tccAACCGCT GCGCTACCA GATTGTCCGC
    601 TGCACCATGG ACGACCATT CCAGATTGCC GAACCCATCC CGGGCATCCG
    651 CTATCAGGCA TTTATCGTGG GTATCCAGGC AGTCAGCCGC AACGGACTTG
    701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGCGGA CGCATTGCGA
35 751 CAAAGCATGG GCGGTCAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA
    801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGGTCGAC CAGACCATCG
    851 CCATCCATTT GGTTCGCGG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
    901 GCCGTAACGG GCGTGGGTTT CGTTTTGGAA GACGACGGCG CGTTCCACTA
    951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
40 1001 AGCCGTTTAC CAATGCCCTT TTGGACAACC AGTCCTACAA AGGCTTCAGT
    1051 ATGCTGCTCG ACATCCGCGA CTCTCCGCGA GGCGAAAAAA CCTTCGACGA
    1101 TTTGTTTATG GATTGGCGG TACGCCTGTC CGGTCACTTG AACCTGAATC
    1151 TGTCAACGA CAAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTA
    1201 CGCACTTATG TATTGGCGCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
45 1251 ACCGGCGGC AAAACGCCC TCGCCTGTT TTCATAA

```

This encodes a protein having amino acid sequence <SEQ ID 530>:

```

50 1 MIYIVLFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
    51 DGKPSGGPVM MPKPQPAVKK PAKQDSAMR NLQEQDAVYI AKQKQAKASP
    101 FKTEIETALE EIGIIGNSAH TVSEPQTGHS APKPADAPAK PVPVPQTPAK
    151 PLITLKELSK VELPWDFVRF DFISYIALTE AKELHALPRL SNRCRYQIVG
    201 CTMDDHFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQADAFa
    251 QSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS
    301 AVTGVGVFLE DDGAFHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFS
    351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
55 401 RTYVLARQSE MLKVGIEPGG KTALRLFS*

```

ORF119ng and ORF119-1 show 98.4% identity over 428 aa overlap:

```

60 orf119ng        MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRD GKPSGGPVM
    |||||:|||||
orf119-1        MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRD GKPSGGSV
    10 20 30 40 50 60

```

-311-

		70	80	90	100	110	120
	orf119ng	MPKPQPAVKKPAKPQDSAMRNLEQDAVYIAKQKQAKASPFKTEIETALEEIGIIGNSAH 					
5	orf119-1	MPKPQPAVKKTAKPQDPAMRNLEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH 					
		70	80	90	100	110	120
		130	140	150	160	170	180
10	orf119ng	TVSEPQTGHSAKPADAPAKPVPVQTPAKPLITLKLKSKVELPWFDVRFDFISYIALTE 					
	orf119-1	TVSEPQTGHSAKPADAPAKPVPVQTPAKPLITLKLKSKVELPWFDVRFDFISYIALTE 					
		130	140	150	160	170	180
		190	200	210	220	230	240
15	orf119ng	AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS 					
	orf119-1	AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS 					
		190	200	210	220	230	240
20		250	260	270	280	290	300
	orf119ng	AFNRQADAFQAQSMGGQTLHTDLAAFIEVASALDAFCARVDQTI A IHLVSPTSISGVELRS 					
	orf119-1	AFNRQVDAFQAQSMGGQTLHTDLAAFIEVASALDAFCARVDQTI A IHLVSPTSISGVELRS 					
25		250	260	270	280	290	300
		310	320	330	340	350	360
	orf119ng	AVTGVGVFLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA 					
30	orf119-1	AVTGVGVFLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA 					
		310	320	330	340	350	360
		370	380	390	400	410	420
	orf119ng	GEKTFDDL FMDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYV LARQSEMLKVGIEPPG 					
35	orf119-1	GEKTFDDL FMDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYV LARQSEMLKVGIEPPG 					
		370	380	390	400	410	420
		429					
40	orf119ng	KTALRLFSX 					
	orf119-1	KTALRLFSX					

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 45 Example 64

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 531>

	1	..GCGCGGCACG	GCACGGAAGA	TTTCTTCATG	AACAACAGCG	ACAC.ATCAG
	51	GCAGATAGTC	GAAAGCACCA	CCGGTACGAT	GAAGCTGCTG	ATTTCCTCCA
50	101	TCGCCCTGAT	TTCATTGGTA	GTCGGCGGCA	TCGGCGTGAT	GAACATCATG
	151	CTGGTGTCCG	TTACCGAGCG	CACCAAAGAA	ATCGGCATAC	GGATGGCAAT
	201	CGGCGCGCGG	CGCGGCAATA	TTTyGCAGCA	GTTTTTGATT	GAGGCGGTGT
	251	TAATCTGCGT	CATCGGCGGT	TTGGTCGGCG	TGGGTTTGTC	CGCCGCCGTC
	301	AGCCTCGTGT	TCAATCATTT	TGTAACCGAC	TCCCGATGG	ACATTTCCGC
	351	CATGTCCGTC	ATCGGCGCGG	TCGCCTGTTC	GACCGGAATC	GGCATCGCGT
55	401	TCGGCTTTAT	GCCTGCCAAT	AAAGCAGCCA	AACTCAATCC	GATAGACGCA
	451	TTGGCACAGG	ATTGA			

This corresponds to the amino acid sequence <SEQ ID 532; ORF134>:

	1	..ARHGTEFFM	NNSDXIRQIV	ESTTGMTKLL	ISSIALISLV	VGGIGVMNIM
	51	LVSVTERKE	IGIRMAIGAR	RGNIXQQFLI	EAVLICVIGG	LVGVGLSAAV
60	101	SLVFNFVTD	FPMDISAMSV	IGAVACSTGI	GIAFGFMPAN	KAAKLNPIDA
	151	LAQD*				

Further work revealed the complete nucleotide sequence <SEQ ID 533>:

```

1  ATGTCGGTGC AAGCAGTATT GGCGCACAAA ATGCGTTCGC TTCTGACGAT
51 GCTCGGCATC ATCATCGGTA TCGCGTCGGT GGTTTCCGTC GTCGCATTGG
101 GCAATGGTTC GCAGAAAAAA ATCCTTGAAG ACATCAGTTC GATAGGGACG
5  151 AACACCATCA GCATCTTCCC GGGGCGCGGC TTCGGCGACA GGCGCAGCGG
201 CAGGATTAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
251 GCTACGTTGC TTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACT
301 TACCGCAACA CCGACCTGAC CGCCTCGCTT TACGGCGTGG GCGAACAATA
351 TTTTCGACGTG CGCGGACTGA AGCTGGAAAC GGGGCGGCTG TTTGACGAAA
10 401 ACGATGTGAA AGAAGACGCG CAGGTCGTCG TCATCGACCA AAATGTCAAA
451 GACAAACTCT TTGCGGACTC GGATCGTTG GGTAAAACCA TTTGTTTTCAG
501 GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAGAC GAAAAACGCTT
551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGGTGATG
601 CACCAAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
15 651 AGACAATGCC AATACCCAGG TTGCCGAAAA AGGGCTGACC GATCTGCTCA
701 AAGCGCGGCA CGGCACGGAA GATTTCCTCA TGAACAACAG CGACACATC
751 AGGCAGATAG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCTCTC
801 CATCGCCCTG ATTTTCATTG TAGTCGGCGG CATCGGCGTG ATGAACATCA
851 TGCTGGTGTC CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA
20 901 ATCGGCGCGC GCGCGGCGAA TATTTTGCAG CAGTTTTTGA TTGAGGCGGT
951 GTTAATCTGC GTCATCGCGG GTTTGGTCGG CGTGGGTTTG TCCGCCGCGC
1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ACTTCCCGAT GGACATTTCC
1051 GCCATGTCCG TCATCGGCGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
1101 GTTCGGCTTT ATGCCTGCCA ATAAAGCAGC CAAACTCAAT CCGATAGACG
25 1151 CATTGGCACA GGATTGA

```

This corresponds to the amino acid sequence <SEQ ID 534; ORF134-1>:

```

1  MSVQAVLAHK MRSLLTMLGI IIGIASVVSVALGNGSQKK ILEDISSIGT
51 NTISIFPGRG FGDRRSGRIK TLTIDDAKII AKQSYVASAT PMTSSGGTTLT
101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
5  151 DKLFADSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
201 HQITGESHTN SITVGIKDNA NTQVAEKGLT DLLKARHGTE DFFMNSDSI
251 RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFPMDIS
351 AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

```

35 Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o648 of *E.coli* (accession number AE000189)

ORF134 and o648 protein show 45% aa identity in 153aa overlap:

```

Orf134: 2  RHGTEDFFMNSDXIRQIVESTTGTMKXXXXXXXXXXXXXVVGIGVMNIMLVSVTERTKEI 61
40  RHG +DFF N D + + VE TT T++ VVGIGVMNIMLVSVTERT+EI
o648: 496 RHGKKDFFTWNMDGVLTVEKTRTLQLFLTLVAVISLVVGIGVMNIMLVSVTERTREI 555

Orf134: 62  GIRMAIGARRGNIXQOFLIEAXXXXXXXXXXXXXXXXXXFNHFVTDFFPMDISAMSVI 121
62  GIRMA+GAR ++ QOFLIEA F+ + + S ++++
45  o648: 556 GIRMAVGARASDVLQOFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPLALL 615

Orf134: 122 GAVACSTGIGIAFGFMPANKAAKLNPIDALAQD 154
122  A CST GI FG++PA AA+L+P+DALA++
o648: 616 LAFLCSTVTGILFGWLPARNAARLDPVDALARE 648

```

50 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF134 shows 98.7% identity over a 154aa overlap with an ORF (ORF134a) from strain A of *N. meningitidis*:

```

55  orf134.pep  ARHGTEFFMNSDXIRQIVESTTGTMKLL
orf134a  GESHTNSITVGIKDNANTQVAEKGLTDLKARHGTEFFMNSDSIRQIVESTTGTMKLL
210 220 230 240 250 260
40 50 60 70 80 90

```



-313-

```

5  orf134.pep  ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG
   orf134a    ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICVIGG
   270      280      290      300      310      320

10  orf134.pep  100      110      120      130      140      150
   orf134a    LVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA
   330      340      350      360      370      380

15  orf134.pep  LAQDX
   orf134a    LAQDX

```

The complete length ORF134a nucleotide sequence <SEQ ID 535> is:

```

20  1  ATGTCGGTGC  AAGCAGTATT  GCGGCACAAA  ATGCGTTTCG  TTCTGACGAT
   51  GCTCGGCATC  ATCATCGGTA  TCGCTTCGGT  TGTCTCCGTC  GTCGCATTGG
  101  GCAACGGTTC  GCAGAAAAAA  ATCCTTGAAG  ACATCAGTTC  GATAGGGACG
  151  AACACCATCA  GCATCTTCCC  AGGGCGCGGC  TTCGGCGACA  GGCGCAGCGG
  201  CAGGATTAAG  ACCCTGACCA  TAGACGACGC  AAAAATCATC  GCCAAACAAA
  251  GCTACGTTGC  TTCCGCCACG  CCCATGACTT  CGAGCGGCGG  CACGCTGACT
  301  TACCGCAATA  CCGACCTGAC  CGCTTCTTTG  TACGGTGTGG  GCGAACAATA
  351  TTTCGACGTG  CGCGGGCTGA  AGCTGGAAAC  GGGGCGGCTG  TTTGACGAAA
  401  ACGATGTGAA  AGAAGACGCG  CAGGTCGTCG  TCATCGACCA  AAATGTCAAA
  451  GACAAACTCT  TTGCGGACTC  GGATCCGTTG  GGTAACCA  TTTGTTCAG
  501  GAAACGCCCC  TTGACCGTCA  TCGGCGTGAT  GAAAAAGAC  GAAAACGCTT
  551  TCGGCAATTC  CGACGTGCTG  ATGCTTTGGT  CGCCCTATAC  GACGGTGATG
  601  CACCAAATCA  CAGGCGAGAG  CCACACCAAC  TCCATCACCG  TCAAAATCAA
  651  AGACAATGCC  AATACCCAGG  TTGCCGAAAA  AGGGCTGACC  GATCTGCTCA
  701  AAGCGCGGCA  CGGCACGAA  GATTTCTTCA  TGAACAACAG  CGACAGCATC
  751  AGGCAGATAG  TCGAAAGCAC  CACCGGTACG  ATGAAGCTGC  TGATTTCTCT
  801  CATCGCCCTG  ATTTTATTGG  TAGTCGGCGG  CATCGGCGTG  ATGAACATCA
  851  TGCTGGTGTC  CGTTACCGAG  CGCACCAAAG  AAATCGGCAT  ACGGATGGCA
  901  ATCGGCGCGC  GGCGCGGCAA  TATTTTGCAG  CAGTTTTTGA  TTGAGGCGGT
  951  GTTAATCTGC  GTCATCGGCG  GTTTGGTTCG  CGTGGGTTTG  TCCGCCGCCG
 1001  TCAGCCTCGT  GTTCAATCAT  TTTGTAACCG  ACTTCCCGAT  GGACATTTCC
 1051  GCCATGTCCG  TCATCGGCGC  GGTGCGCTGT  TCGACCGGAA  TCGGCATCGC
 1101  GTTCGGCTTT  ATGCTGCGCA  ATAAAGCAGC  CAAACTCAAT  CCGATAGATG
 1151  CATTGGCGCA  GGATTGA

```

This encodes a protein having amino acid sequence <SEQ ID 536>:

```

45  1  MSVQAVLAHK  MRSLLTMLGI  IIGIASVVS  VALNGSQKK  ILEDISSIGT
   51  NTISIFPGRG  FGDRRSGRIK  TLTIDDAKII  AKQSYVASAT  PMTSSGGTLT
  101  YRNTDLTASL  YGVGEQYFDV  RGLKLETGRL  FDENDVKEDA  QVVVIDQNVK
  151  DKLFADSDPL  GKTILFRKRP  LTVIGVMKKD  ENAFGNSDVL  MLWSPYTTVM
  201  HQITGESHTN  SITVKIKDNA  NTQVAEKGLT  DLLKARHGTE  DFFMNSDSI
  251  RQIVESTTGT  MKLLISSIAL  ISLVVGGIGV  MNIMLVSVTE  RTKEIGIRMA
  301  IGARRGNILQ  QFLIEAVLIC  VIGGLVGVGL  SAAVSLVFNH  FVTDFPMDIS
  351  AMSVIGAVAC  STGIGIAFGF  MPANKAAKLN  PIDALAQD*

```

50 ORF134a and ORF134-1 show 100.0% identity in 388 aa overlap:

```

55  orf134a.pep  MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGSQKKILEDISSIGTNTISIFPGRG
   orf134-1    MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGSQKKILEDISSIGTNTISIFPGRG

  orf134a.pep  FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
   orf134-1    FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV

  orf134a.pep  RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
   orf134-1    RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD

  orf134a.pep  ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLKARHGTE
   orf134-1    ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLKARHGTE

```

5 orf134a.pep DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA  
 orf134-1 DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA  
 orf134a.pep IGARRGNILQQFLIEAVLICVIGGLVGVLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC  
 orf134-1 IGARRGNILQQFLIEAVLICVIGGLVGVLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC  
 10 orf134a.pep STGIGIAFGFMPANKAAKLNPIDALAQDX  
 orf134-1 STGIGIAFGFMPANKAAKLNPIDALAQDX

### Homology with a predicted ORF from *N.gonorrhoeae*

15 ORF134 shows 96.8% identity over a 154aa overlap with a predicted ORF (ORF134.ng) from *N.gonorrhoeae*:

orf134.pep ARHGTEDFFMNSDXIRQIVESTTGTMKLL 30  
 orf134ng GESHTNSITVKIKDNANTRVAEKGLAELLKARHGTEDFFMNSDSIRQMVESTTGTMKLL 264  
 orf134.pep ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG 90  
 orf134ng ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICIIGG 324  
 25 orf134.pep LVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA 150  
 orf134ng LVGVGLSAAVSLVFNHFVTDFFPMDISAASVIGAVACSTGIGIAFGFMPANKAAKLNPIDA 384  
 30 orf134.pep LAQD 154  
 orf134ng LAQD 388

The complete length ORF134ng nucleotide sequence <SEQ ID 537> is:

1 ATGTCGGTGC AAGCAGTATT GGCGCACAAA ATGCGTTCGC TTCTGACCAT  
 35 51 GCTCGGCATC ATCATCGGTA TCGCTTCGGT TGTCTCCGTC GTCGCGCTGG  
 101 GCAACGGTTC GCAGAAAAAA ATCCTCGAAG ACATCAGTTC GATGGGGACG  
 151 AACACCATCA GCATCTTCCC CGGGCGCGGC TTCGGCGACA GGCGCAGCGG  
 201 CAAAATCAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA  
 251 GCTACGTTGC CTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACC  
 301 TACCGCAATA CCGACCTGAC CGCTTCTTTG TACGGTGTGG GCGAACAATA  
 351 TTTTCGCGTG CGCGGGCTGA AGCTGGAAC GGGGCGGCTG TTTGATGAGA  
 401 ACGATGTGAA AGAAGACGCG CAAGTCGTCG TCATCGACCA AAATGTCAAA  
 451 GACAAACTCT TTGCGGACTC GGATCCGTTG GGTAAAACCA TTTTGTTCAG  
 501 GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAAGAC GAAAACGCTT  
 551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGGTGATG  
 45 601 CACCAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA  
 651 AGACAATGCC AATACCCGGG TTGCCGAAAA AGGGCTGGCC GAGCTGCTCA  
 701 AAGCAGGCA CGGCACGGAA GACTTCTTTA TGAACAACAG CGACAGCATC  
 751 AGGCAGATGG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCCTC  
 801 CATCGCCCTG ATTTCAATGG TAGTCGGCGG CATCGGTGTG ATGAACATTA  
 50 851 TGCTGGTGTC CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA  
 901 ATCGGCGCGC GCGCGGCAA TATTTTGCAG CAGTTTTTGA TTGAGGCGGT  
 951 GTTAATCTGC ATCATCGGAG GCTTGGTCGG CGTAGGTTTG TCCGCCGCCG  
 1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ATTTCCCGAT GGACATTTCC  
 1051 GCGGCATCCG TTATCGGGG GGTGCGCTGT TCGACCGGAA TCGGCATCGC  
 55 1101 GTTCGGCTTT ATGCTTGCCA ATAAGGCAGC CAAACTCAAT CCGATAGATG  
 1151 CATTGGCGCA GGATTGA

This encodes a protein having amino acid sequence <SEQ ID 538>:

1 MSVQAVLAHK MRSLLTMLGI IIGIASVVSVALGNGSQKK ILEDISSMGT  
 51 NTISIFPGRG FGDRRSGKIK TLTIDDAKII AKQSYVASAT PMTSSGGTILT  
 60 101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK  
 151 DKLFADSDPL GKILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM  
 201 HQITGESHTN SITVKIKDNA NTRVAEKGLA ELLKARHGT DFFMNSDSI  
 251 RQMVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA  
 301 IGARRGNILQ QFLIEAVLIC IIGGLVGVGL SAAVSLVFNH FVTDFPMDIS

-315-

351 AASVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD\*

ORF134ng and ORF134-1 show 97.9% identity in 388 aa overlap:

```

5   orf134ng      MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSMGNTNISIFPGRG
    orf134-1      MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSIGTNTNISIFPGRG

10  orf134ng      FGDRRSKGIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
    orf134-1      FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV

15  orf134ng      RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
    orf134-1      RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD

20  orf134ng      ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGTE
    orf134-1      ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTE

25  orf134ng      DFFMNSDSIRQMVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
    orf134-1      DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA

    orf134ng      IGARRGNILQQFLIEAVLICIIIGGLVGVGLSAAVSLVFNHFVTDFFPMDISAASVIGAVAC
    orf134-1      IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC

    orf134ng      STGIGIAFGFMPANKAAKLNPIDALAQDX
    orf134-1      STGIGIAFGFMPANKAAKLNPIDALAQDX

```

30 ORF134ng also shows homology to an *E. coli* ABC transporter:

```

sp|P75831|YBJZ_ECOLI_HYPOTHETICAL_ABC_TRANSPORTER_ATP-BINDING_PROTEIN_YBJZ >gi5
(AE000189) o648; similar to YBBA_HAEIN SW: P45247 [Escherichia coli] Length =
648
Score = 297 bits (753), Expect = 6e-80
Identities = 162/389 (41%), Positives = 230/389 (58%), Gaps = 1/389 (0%)

35  Query: 1  MSVQAVLAHKMRSLLTMLXXXXXXXXXXXXXXXXXVVGIGVMNIMLVSVTERTKEIGIRM 60
    M+ +A+ A+KMR+LLTML +G+ +++ +L DI S+GTNTI ++PG+
    Sbjct: 260 MAWRALAANKMRTLTLTMLGIIIGIASVVSIVVVGDAKQMVLAIRSIGTNTIDVYPGKD 319

40  Query: 61 FGDRRSKGIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV 120
    FGD + L DD I KQ +VASATP S L Y N D+ AS GV YF+V
    Sbjct: 320 FGDDDPQYQQALKYDDLIAIQKQPWVASATPAVSQNLRLRYNNVDVAASANGVSGDYFNV 379

45  Query: 121 RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFAD-SDPLGKTILFRKRPLTVIGVMKK 179
    G+ G F++ + AQVVV+D N + +LF +D +G+ IL P VIGV ++
    Sbjct: 380 YGMTFSEGNTFNQEQLNGRAQVVVLDNTRRQLFPHKADVVGIVLVGNMPARVIGVAEE 439

50  Query: 180 DENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGT 239
    ++ FG+S VL +W PY+T+ ++ G+S NSITV++K+ ++ AE+ L LL RHG
    Sbjct: 440 KQSMFGSSKVLRVLPYSTMSGVRMGQSWLNSITVRVKEGFDSAEAEQQLTRLLSLRHGK 499

55  Query: 240 EDFFMNSDSIRQMVESTTGTMKXXXXXXXXXXXXXVVGIGVMNIMLVSVTERTKEIGIRM 299
    +DFF N D + + VE TT T++ VVGIGVMNIMLVSVTERT+EIGIRM
    Sbjct: 500 KDFFTWNMDGVLKTVKTTRTLQLFLTLVAVISLVVGGIGVMNIMLVSVTERTREIGIRM 559

60  Query: 300 AIGARRGNILQQFLIEXXXXXXXXXXXXXXXXXXXXXFNHFVTDFFPMDISAASVIGAVA 359
    A+GAR ++LQQFLIE F+ + + S +++ A
    Sbjct: 560 AVGARASDVLQQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPLALLLAFL 619

    Query: 360 CSTGIGIAFGFMPANKAAKLNPIDALAQD 388
    CST GI FG++PA AA+L+P+DALA++
    Sbjct: 620 CSTVTGILFGWLPARNAARLDPVDALARE 648

```

Based on this analysis, including the presence of the leader peptide and transmembrane regions in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 65

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 539>:

```

1  ..GGGACGGGAG CGATGCTGCT GCTGTTTAC GCGGTAACGA T.CTGCCTTT
51  GGCCACTGGC GTTACCCTGA GTTACACCTC GTCGATTTTT TTGGCGGTAT
101 TTTCTTCCT GATTTTGAAA GAACGGATT CCGTTTACAC GCAGCGGTG
151 CTGCTCCTTG GTTTTGCCGG CGTGGTATTG CTGCTTAATC CCTCGTTCCG
10  201 CAGCGGTCAG GAAACGGCGG CACTCGCCGG GCTGGCGGGC GGCGCGATGT
251 CCGGCTGGGC GTATTTGAAA GTGCGCGAAC TGTCTTTGGC GGGCGAAGCC
301 GGCTGGCGCG TCGTGTTTTA CTTTCCGTG ACAGGTGTGG CGATGTCGTC
351 GGTTTGGGCG ACGCTGACCG GCTGGCACAC CCTGTCCTTT CCATCGGCAG
401 TTTATCTGTC GTGCATCGGC GTGTCCGCGC TGATTGCCCA ACTGTCGATG
15  451 ACGCGCGCCT ACAAAGTCGG CGACAAATC ACGGTGCGCT CGCTTCTCTA
501 TATGACCGTC GTTTTTTCCG CTCTGTCTGC CGCATTTTTT CTGGGCGAAG
551 AGCTTTTCTG GCAGGAAATA CTCGGTATGT GCATCATCAT CCTCAGCGGT
601 ATTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 540; ORF135>:

```

20  1  ..GTGAMLLFY AVTILPLATG VTLSYTSSIF LAVFSFLILK ERISVYTQAV
51  LLLGFAGVVL LLNPSFRSQ ETAALAGLAG GAMSGWAYLK VRELSLAGEP
101 GWRVVFYLSV TGVAMSSVWA TLTGWHLSF PSAVYLSCIG VSALIAQLSM
151 TRAYKVGDKF TVASLSYMTV VFSALSAAFF LGEELFWQEI LGMCIISAV
201 F*

```

25 Further work revealed the complete nucleotide sequence <SEQ ID 541>:

```

1  ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC
51  GCGGCCTGC TTTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAA
101 AATTTGCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA
151 ACCGTTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA mCTTCCGCAC
30  201 GCCCCATTGG AAAAACCCT TAAACCGCAG TATGGTCGGG ACGGGGCGGA
251 TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTTGGC CACTGGCGTT
301 ACCCTGAGTT ACACCTCGTC GATTTTTTTG GCGGTATTTT CCTTCCTGAT
351 TTTGAAAGAA CGGATTTCCG TTTACACGCA GGCGGTGCTG CTCCTTGGTT
401 TTGCCGGCGT GGTATTGCTG CTTAATCCCT CGTTCCGCAG CGGTGAGGAA
35  451 ACGGCGGCAC TCGCCGGGCT GGCGGGCGGC GCGATGTCCG GCTGGGCGTA
501 TTTGAAAGTG CGCGAACTGT CTTTGGCGGG CGAACCCGGC TGGCGCGTCG
551 TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTCGTGCGT TTGGGCGACG
601 CTGACCGGCT GGCACACCCT GTCCTTTCCA TCGGCAGTTT ATCTGTCGTG
651 CATCGGCGTG TCCGCGCTGA TTGCCCACT GTCGATGACG CGCGCTACA
40  701 AAGTCGGCGA CAAATTCACG GTTGCTCGC TTTCTTATAT GACCGTCGTT
751 TTTTCCGCTC TGTCTGCCG ATTTTTTCTG GGCGAAGAGC TTTTCTGGCA
801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
851 TCCGCCCCAC TGCCTTCAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA
901 TAA

```

45 This corresponds to the amino acid sequence <SEQ ID 542; ORF135-1>:

```

1  MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRLFS
51  TVALGAAVL RDXFRTPHW KNHLNRSVMG TGAMLLLFYA VTHLPLATGV
101 TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVLL LNPSFRSQE
151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVMSSVWAT
50  201 LTGWHLSFP SAVYLSCIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
251 FSALSAAFFL GEELFWQEIL GMCIILSGI LSSIRPTAFK QRLQSLFRQR
301 *

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF135 shows 99.0% identity over a 197aa overlap with an ORF (ORF135a) from strain A of *N. meningitidis*:

```

5      orf135.pep      10      20      30
                        GTGAMLLLFYAVTILPLATGVTLSTSSIF
                        |||
      orf135a      STVALGAAVLRDTRTPHWKNHLNRSMTGAMLLLFYAVTHLPLATGVTLSTSSIF
                        50      60      70      80      90      100

10     orf135.pep      40      50      60      70      80      90
                        LAVFSFLILKERISVYTQAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK
                        |||
      orf135a      LAVFSFLILKERISVYTQAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK
                        110     120     130     140     150     160

15     orf135.pep      100     110     120     130     140     150
                        VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSLSCIGVSALIAQLSM
                        |||
      orf135a      VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSLSCIGVSALIAQLSM
                        170     180     190     200     210     220

20     orf135.pep      160     170     180     190     200
                        TRAYKVGDKFTVASLSYMTVVFSAALAEELFWQEILGMCIIISAVFX
                        |||
      orf135a      TRAYKVGDKFTVASLSYMTVVFSAALAEELFWQEILGMCIIISAVFX
                        230     240     250     260     270     280

      orf135a      KQRLQSLFRQRX
                        290     300

```

30 The complete length ORF135a nucleotide sequence <SEQ ID 543> is:

```

      1  ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC
      51  GCGGCGCTGC TTTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAA
     101  AATTTGCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA
     151  ACCGTTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA CCTTCCGCAC
     201  GCCCCATTGG AAAAACCAC TAAACCGCAG TATGGTCGGG ACGGGGGCGA
     251  TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTTGGC CACCGGCGTT
     301  ACCCTGAGTT ACACCTCGTC GATTTTTTTG GCGGTATTTT CCTTCCTGAT
     351  TTTGAAAGAA CGGATTTCGG TTTACACGCA GCGCGTGCTG CTCCTTGCTT
     401  TTGCCGGCGT GGTATTGCTG CTTAATCCCT CGTTCGCGAG CGGTCAGGAA
     451  ACGGCGGCAC TCGCCGGGCT GCGGGGCGGC GCGATGTCCG GCTGGGCGTA
     501  TTTGAAAGTG CGCGAACTGT CTTTGGCGGG CGAACCGGCG TGGCGCGTCG
     551  TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTCATCGGT TTGGGCGACG
     601  CTGACCGGCT GGCACACCTT GTCCTTTCCA TCGGCAGTTT ATCTGTCTGT
     651  CATCGGCGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCCTACA
     701  AAGTCGCGCA CAAATTCACG GTTGCCTCGC TTTCCTATAT GACCGTCGTT
     751  TTTTCCGCTC TGTCTGCCGC ATTTTCTCTG GCCGAAGAGC TTTTCTGGCA
     801  GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
     851  TCCGCCACAC TGCCTCAAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA
     901  TAA

```

50 This encodes a protein having amino acid sequence <SEQ ID 544>:

```

      1  MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMFLS
      51  TVALGAAAVL RRDTRTPHW KNHLNRSMTG TGAMLLLFYA VTHLPLATGV
     101  TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVLL LNPSFRSGQE
     151  TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVAMSSVWAT
     201  LTGWHTLSFP SAVYLSLSCIGV SALIAQLSMT RAYKVGDKFT VASLSYMTV
     251  FSALSAAFFL AEELFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR
     301  *

```

ORF135a and ORF135-1 show 99.3% identity in 300 aa overlap:

```

60     orf135a.pep      MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL
                        |||
      orf135-1      MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL

```

```

5  orf135a.pep  RRDTFRTPHWKNHLNRSMVGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE
   orf135-1    RRDXFRTPHWKNHLNRSMVGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE

10 orf135a.pep  RISVYTQAVLLLGAGVLLNPSFRSQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
   orf135-1    RISVYTQAVLLLGAGVLLNPSFRSQETAALAGLAGGAMSGWAYLKVRELSLAGEPG

15 orf135a.pep  WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
   orf135-1    WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT

   orf135a.pep  VASLSYMTVVFSALSAAFFLAEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR
   orf135-1    VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR

```

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF135 shows 97% identity over a 201aa overlap with a predicted ORF (ORF135ng) from

20 *N.gonorrhoeae*:

```

   orf135.pep  GTGAMLLLFYAVTXLPLATGVTLSTSSIF  30
   orf135ng    STVTLGAAAVLRRDTFRTPHWKNHLNRSMVGTGAMLLLFYAVTHLPLTTGVTLSTSSIF  335

25 orf135.pep  LAVFSFLILKERISVYTQAVLLLGAGVLLNPSFRSQETAALAGLAGGAMSGWAYLK  90
   orf135ng    LAVFSFLILKERISVYTQAVLLLGAGVLLNPSFRSQEPAALAGLAGGAMSGWAYLK  395

30 orf135.pep  VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSM  150
   orf135ng    VRELSLAGEPGWRVVFYLSATGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSM  455

   orf135.pep  TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAVF  201
   orf135ng    TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAAF  506

```

An ORF135ng nucleotide sequence <SEQ ID 545> was predicted to encode a protein having amino acid sequence <SEQ ID 546>:

```

40  1  MPSEKAFFRRH  LRTASFOGLH  LHHFHQKVKGK  CGIIGFGIHI  FPTLLPAAQG
   51  ILDIQLGLFR  IDFAALAVYR  RTQVDFIHTV  IDGIASDQAF  SEVVQILRRL
  101  NLGHFTDTHL  IAQARRFIAD  FGNIRPMRRG  EAKTFCRCFR  FDGIDGIHGD
  151  FRQCGHINRL  APGKDCRNGK  RDKVFFHTRH  YNQVCLEKTN  CSARKIKFRH
  201  QKQAKTHSTS  LAARFTIRPS  LSQRPFMDTA  KKDILGSGWM  LVAAACFTVM
  251  NVLIKEASAK  FALGSGELVF  WRMLFSTVTL  GAAAVLRRDT  FRTPHWKNHL
  301  NRSMVGTGAM  LLLFYAVTHL  PLTTGVTLST  TSSIFLAVFS  FLILKERISV
45  351  YTQAVLLLG  AGVLLLNPS  FRSGQEPAL  AGLAGGAMSG  WAYLKVRELS
   401  LAGEPGWRVV  FYLSATGVAM  SSVWATLTGW  HTLSFPSAVY  LSGIGVSALI
   451  AQLSMTRAYK  VGDKFTVASL  SYMTVVFSAL  SAAFFLGEEL  FWQEILGMC
   501  IISAAF*

```

Further work revealed the following gonococcal sequence <SEQ ID 547>:

```

50  1  ATGGATACCG  CAAAAAAGA  CATTTTAGGA  TCGGGCTGGA  TGCTGGTGGC
   51  GGCGGCCTGC  TTCACCGTTA  TGAACGTATT  GATTAAAGAG  GCATCGGCAA
  101  AATTTGCCCT  CGGCAGCGGC  GAATTGGTCT  TTTGGCGCAT  GCTGTTTTCA
  151  ACCGTTACGC  TCGGTGCTGC  CGCCGTATTG  CGGCGCGACA  CCTTCCGCAC
  201  GCCCATTTGG  AAAAACCAC  TAAACCGCAG  TATGGTCGGG  ACGGGGCGA
55  251  TGCTGCTGCT  GTTTACGCG  GTAACGCATC  TGCCTTGAC  AACCGCGGTT
   301  ACCTGAGTT  ACACCTCGTC  GATTTTttg  GCGGTATTTT  CCTTCCTGAT
   351  TTTGAAAGAA  CGGATTTCCG  TTTACACGCA  GGCGGTGCTG  CTCCTTGGTT
   401  TTGCCGGCGT  GGTATTGCTG  CTTAATCCCT  CGTTCCGCAG  CGGTACAGAA
   451  CCGCGGCAC  TCGCCGGGCT  GGCGGGCGGC  GCGATGTCCG  GCTGGGCGTA
60  501  TTTGAAAGTG  CGCGAACTGT  CTTTGGCGGG  CGAACCGGCG  TGGCGGCTCG
   551  TGTTTACCT  TTCCGCAACC  GGCGTGGCGA  TGTCGTGcgt  ttgggcgacg
   601  Ctgaccggct  ggCACaccct  GTCCTTTcca  tcggcagttt  ATCtgtCGGG

```

-319-

5 651 CATCGGCGTG tccgcgCtgA TTGCCCAaCT GtcgatgAcg cGCGcctaca  
 701 aaGTCGGCGA CAAATTCACG GTTGCTCGC tttcctaTat gaccgtcGTC  
 751 TTTTCGCCCC TGTCTGCCGC ATTTTTTCTg ggcgaagagc ttttCtggCA  
 801 GGAAATACTC GGTATGTGCA TCATTatcct CAGCGGCATT TTGAGCAGCA  
 851 TCCGCCCCAT TGCCTTCAA CAGCGGCTGC AAGCCCTCTT CCGCCAAAGA  
 901 TAA

This corresponds to the amino acid sequence <SEQ ID 548; ORF135ng-1>:

10 1 MDTAKKDILG SGWMLVAAAC FTVMNVLIKE ASAKFALGSG ELVFWRMLFS  
 51 TVTLGAAAVL RRDFTFPHW KNHLNRSVMG TGAMLLLFYA VTHLPLTTGV  
 101 TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVLL LNPSFRSGQE  
 151 PAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSAT GVAMSSVWAT  
 201 LTGWHTLSFP SAVYLSGIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV  
 251 FSALSAAFFL GEELFWQEIL GMCIIILSGI LSSIRPIAFK QRLQALFRQR  
 301 \*

15 ORF135ng-1 and ORF135-1 show 97.0% identity in 300 aa overlap:

orfl35ng-1.pep MDTAKKDILGSGWMLVAAACFTVMNVLIKEASAKFALGSGELVFWRMLFSTVTLGAAAVL  
 orfl35-1 MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMLFSTVALGAAAVL  
 20 orfl35ng-1.pep RRDFTFPHWKNHLNRSVMGTGAMLLLFYAVTHLPLTTGVTLSTSSIFLAVFSFLILKE  
 orfl35-1 RRDXFRTPHWKNHLNRSVMGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE  
 25 orfl35ng-1.pep RISVYTQAVLLLGFGVLLLNPSFRSGQEPALAGLAGGAMSGWAYLKVRELSLAGEPG  
 orfl35-1 RISVYTQAVLLLGFGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG  
 orfl35ng-1.pep WRVVFYLSATGVAMSSVWATLTGWHTLSFSAVYLSGIGVSALIAQLSMT RAYKVGDKFT  
 30 orfl35-1 WRVVFYLSVTGVAMSSVWATLTGWHTLSFSAVYLSGIGVSALIAQLSMT RAYKVGDKFT  
 orfl35ng-1.pep VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPIAFKQRLQALFRQR  
 orfl35-1 VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR

35 Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 66

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 549>:

40 1 ATGAAGCGGC GTATAGCCGT CTTCGTCCTG TTCCCGCAGA TAATCCGAGT  
 51 TTTGGGACAA CTGTTGCCGA AAATCGTCAA TACAGTTCCG GCACATCGGA  
 101 TGCTCTTCCA GATTTTCGGG ATGTTCTTTT TCTTCATACA CCAGCAATAT  
 151 CTGCCCGGGA TCGCCGAAAT CGATTCCCCA TGCGGCATCG TGTTCGGTGC  
 45 201 GCTCCTCTTC CGTCATCTGC CCGCGCATTG CCTGTATGGT AAAGCCGCCG  
 251 TAGGGGATGC CgTTGCACAC GAACATCCAG TCGTGTATGT CGTCAACCGG  
 301 AACGCAAACG cTTTCGCCTT GTTCGACATT GGTCAGTTCG CCsGGTTCAT  
 351 TGTTACGAC ACCGTAAAT TAAAGACCGT CAAAATAAAT ATCGTCGATC  
 401 CACATATGTT CGCAAATTTC GCCGTCTTCG CCGTCTTGA AAAAAGGGAC  
 451 TTTGACCATG GCAAATCCA AGGCGGAAAT AATGCGGCGG CGTTCCCAAA  
 50 501 AAAGcTCGCG CCAAAAATAT TTGAATGTTT TACGGGCGCG TTCGTGCGCA  
 551 CGGTTTACCG GTTCGTCTGC CTGTTCTACA TAATAAATGA CGGAATCGCC  
 601 CATCATATCT GTCCTCAAC GTGTACGGTA TCTGTTTGCA CCTTACTGCG  
 651 GCTTTCTgC kTCGGCATCC GATTCCGATT TGAAAAGTTC mmrwyATTCTG  
 701 GAATAG

55 This corresponds to the amino acid sequence <SEQ ID 550; ORF136>:

1 MKRRIAEFVL FPQIIRVLGQ LLPKIVNTVP AHRMLFQIFG MFFFFIHQQY  
 51 LPGIAEIDSP CGIVFGALLF RHLPAHCLYG KAAVGDAVAH EHPVADVNR

101 NANAFALEFDI GQFAXFIVQH TVNIKTVKIN IVDPHMFANF AVFAVLEKRD  
 151 FDHGKIQQGN NAAAFPKKLA PKIFECFTGA FVGTVYRFVC LFYIINDGIA  
 201 HHSAPQVRVY LFAPYCGFLP SASDSLKSS XXSE\*

Further work revealed the complete nucleotide sequence <SEQ ID 551>:

5 1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGTTCCCGC AGATAATCCG  
 51 AGTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC  
 101 GGATGCTCTT CCAGATTTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA  
 151 TATCTGCCCC GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCGG  
 201 TCGCTCCTC TTCCGTCATC TGCCCGCGCA TTGCCTGTAT GGTAAAGCCG  
 10 CCGTAGGGGA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCGTCAAC  
 301 CGGAACGCAA ACGCTTTCGC CTTGTTTCGAC ATTGGTCAGT TCGCCGGGTT  
 351 CATGTGTCAG CACACCGTAA ATATAAAGAC CGTCAAAAATA AATATCGTCG  
 401 ATCCACATAT GTTCGCAAAT TTCGCCGTCT TCGCCGTCTT GGAAAAAGG  
 451 GACTTTGACC ATGGCAAAAT CCAAGGCGGA AATAATGCGG CGGCGTTCCTC  
 15 501 AAAAAAGCTC GCGCAAAAAA TATTTGAATG TTTTACGGGC GCGTTCGTCTG  
 551 GCACGGTTTA CCGGTTTCGTC TGCCTGTTCT ACATAATAAA TGACGGAATC  
 601 GCCCATCATT CTGCTCCTCA ACGTGTACGG TATCTGTTG CACCTTACTG  
 651 CGGCTTTCTG CCTTCGGCAT CCGATTCCGA TTTGAAAAGT TCCAAATATT  
 701 CGGAATAG

20 This corresponds to the amino acid sequence <SEQ ID 552; ORF136-1>:

1 MMKRRIAVFV LFPQIIRVLG QLLPKIVNTV PAHRMLFQIF GMFFFFIHQQ  
 51 YLPGIAEIDS PCGIVFGALL FRHLPAHCLY GKAAVGDAVA HEHPVADVNN  
 101 RNANAFALFD IGQFAGFIVQ HTVNIKTVKI NIVDPHMFAN FAVFAVLEKR  
 151 DFDHGKIQQG NNAAFPKKL APKIFECFTG AFVGTVYRFV CLFYIINDGI  
 25 201 AHHSAPQVRV YLFAPYCGFL PSASDSLKS SKYSE\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF136 shows 71.7% identity over a 237aa overlap with an ORF (ORF136a) from strain A of *N. meningitidis*:

30 orf136.pep 10 20 30 40 50 59  
 MKRRIAVFVLFPPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGIAEIDS  
 orf136a 10 20 30 40 50 60  
 MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFIHQQYLPGIAEIDS  
 35 60 70 80 90 100 110 119  
 orf136.pep PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVNNRNANAFALFDIGQFAXFIVQ  
 orf136a PCGIVFGTLLFRHXSTHCLYGKAAVGNAVAHEHPVADVNNRNANAFALFDIGQFAGFIVQ  
 40 70 80 90 100 110 120  
 120 130 140 150 160 170 179  
 orf136.pep HTVNIKTVKINIVDPHMFANFAVFAVLEKRD FDHGKIQQGNNAAFPKKLAPKIFECFTG  
 45 orf136a HAINVKTVKINIVDPHMFANFAVFAVLEKRALMAKSKXXMRRRSQKSSRQKYNVLR  
 130 140 150 160 170 180  
 180 190 200 210 220 230  
 50 orf136.pep AFVGTVYRFVCLFYIINDGIAHH---SAPQVRVRYLFAPYCGFLPSASDSLKSSXXSEX  
 orf136a R---SPARFTGLSACSTXMTESPIISAPQVRVRYLFAPYCGFLPSASDSLKSSKYSEX  
 190 200 210 220 230

The complete length ORF136a nucleotide sequence <SEQ ID 553> is:

55 1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGCTCATGC AGAAAATCCG  
 51 GATTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC  
 101 GGATGCTCTT CCAGATNTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA  
 151 TACCTGCCCC GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCGG  
 201 TACGCTCCTC TTCCGTCATC NGTCCACGCA TTGCCTGTAT GGTAAAGCCG  
 251 CCGTAGGGAA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCGTCAAC



-321-

5  
301 CGGAACGCAA ACGCTTTCGC CTTGTTTCGAC ATTGGTCAGT TCGCCGGGTT  
351 CATTGTTTCAG CACGCCATAA ATGTAAAGAC CGTCAAATAA AATATCGTCG  
401 ATCCACATAT GTTCGCAAAAT TTCGCCNTCT TCGCCGTCCT GGAAAAAAGG  
451 GCTTTGACCA TGGCAAAATC TAAGGNGNNA NNGATGCGGC GCGGTTCCCA  
501 AAAAAGCTCG CGCCAAAAAT ATTGAATGT TTTGCGGGCG CGTTCGCGCG  
551 CACGGTTTAC CGGTTTGTCT GCCTGTTCTA CATAATAAAT GACGGAATCG  
601 CCCATCATAT CTGCTCCTCA ACGTGTACGG TATCTGTTTG CACCTTACTG  
651 CGGCTTTCTG CCTTCGGCAT CCGATTGCGA TTTGAAAAGT TCCAAATATT  
701 CGGAATAG

10 This encodes a protein having amino acid sequence <SEQ ID 554>:

1 MMKRRIAVFV LLMQKIRILG QLLPKIVNTV PAHRMLFQXF GMFFFFFIHQQ  
51 YLPGIAEIDS PCGIVFGTLL FRHXSTHCLY GKAAVGNVA HEHPVADVNN  
101 RNANAFALFD IGQFAGFIVQ HAINVKT VKI NIVDPHMFAN FAXFAVLEKR  
151 ALTMAKSKXX XMRRRSQKSS RQKYNLVLRA RSPARFTGLS ACST\*\*MTES  
15 201 PIISAPQVRV YLFAPYCGFL PSASDSLKS SKYSE\*

ORF136a and ORF136-1 show 73.1% identity in 238 aa overlap:

20  
25  
30  
35  
40

orf136a.pep	10	20	30	40	50	60
orf136-1	10	20	30	40	50	60
orf136a.pep	70	80	90	100	110	120
orf136-1	70	80	90	100	110	120
orf136a.pep	130	140	150	160	170	180
orf136-1	130	140	150	160	170	180
orf136a.pep	190	200	210	220	230	
orf136-1	190	200	210	220	230	

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF136 shows 92.3% identity over a 234aa overlap with a predicted ORF (ORF136ng) from *N.gonorrhoeae*:

45  
50  
55

orf136.pep	MKRRIAVFVLFQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFFIHQQYLPGIAEIDS	59
orf136ng	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFFIHRQYLPGIAEIDS	60
orf136.pep	PCGIVFGALLFRHLPAHCLYGKAAVGDVAHEHPVADVNNRANAFALFDIGQFAXFIVQ	119
orf136ng	PGGIVFGTLLFRHLSAHCLYGKAAVGDVAHEHPVADVANNRANAFALFDIGQSAGFIVQ	120
orf136.pep	HTVNIKT VKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAFPKKLAPKIFECFTG	179
orf136ng	HTVNIKT VKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAFPKKLAPKVFECFTG	180
orf136.pep	AFVGTVYRFVCLFYIINDGIAHHSAPQVRVRYLFAPYCGFLPSASDSLKSXXSE	234
orf136ng	AFAGTVYRFVCLFYIINDGIAHHTAPQVRVRYLFAPYRGFLPPASDSLKSXYSE	235

The complete length ORF136ng nucleotide sequence <SEQ ID 555> is:

60  
1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGCTCATGC AGAAAATCCG  
51 GATTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC

-322-

101 GGATGCTCTT CCAAATTTTC GGGATGTTCT TTTTCTTCAT ACACCGGCAA  
 151 TACCTGCCCG GGATCGCCGA AATCGATTCC CCAGGCGGTA TCGTGTTCGG  
 201 TACGCTCCTC TTCCGTCATC TGTCCGCGCA TTGCCTGTAC GGTAAGCCG  
 251 CCGTAGGGGA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCGCCAAC  
 5 301 CGGAACGCAA ACGCTTTTCG CTTGTTTCGAC ATTGGTCAGT CCGCCGGGTT  
 351 CATTTGTTT CACACCGTAA ATATAAGAC CGTCAAAATA AATATCGTCG  
 401 ATCCACATAT GTTCGCAAT TTCGCCGTCT TCGCCGTCTT GGAAAAAAGG  
 451 GACTTTGACC ATGGCAAAAT CCAAGGCGGA AATAATGCGG CGGCGTCCCC  
 501 AAAAAAGCTC GCGCCAAAAG TATTTGAATG TTTTACGGGC GCGTTCGCGC  
 10 551 GCACGGTTTA CCGGTTTCGT TGCCTGTTCT ACATAATAA TGACGGAATC  
 601 GCCCATCATA CTGCTCCTCA ACGTGTACGG TATCTGTTG CACCTTACCG  
 651 CCGTTTTCTA CCTCCGGCAT CCGATTCCGA TTTGAAAAGT TCCAAATATT  
 701 CGGAATAG

This encodes a protein having amino acid sequence <SEQ ID 556>:

15 1 MMKRRIAVFV LLMQKIRILG QLLPKIVNTV PAHRMLFQIF GMFFFFIHRQ  
 51 YLPGIAEIDS PGGIVFGTLL FRHLSAHCLY GKAAGDAVA HEHPVADVAN  
 101 RNANAFALFD IGQSAGFIVQ HTVNIKT VKI NIVDPHMFAN FAVFAVLEKR  
 151 DFDHGKIQQG NNAAPFKKL APKVFECFTG AFAGTVYRFV CLFYIINDGI  
 201 AHHTAPQVRV YLFAPYRGFL PPASDSLKS SKYSE\*

20 ORF136ng and ORF136-1 show 93.6% identity in 235 aa overlap:

orf136ng MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHRQYLPGLAEIDS  
 orf136-1 MMKRRIAVFVLPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHRQYLPGLAEIDS  
 25 orf136ng PGGIVFGTLLFRHLSAHCLYGKAAGDAVAHEHPVADVANRNANAFALFDIGQSAGFIVQ  
 orf136-1 PCGIVFGALLFRHLPAHCLYGKAAGDAVAHEHPVADVNNANAFALFDIGQFAGFIVQ  
 30 orf136ng HTVNIKT VKINIVDPHMFANFAVFAVLEKRDFDHGKIQQGNNAAAFPKKLAPKVFECFTG  
 orf136-1 HTVNIKT VKINIVDPHMFANFAVFAVLEKRDFDHGKIQQGNNAAAFPKKLAPKIFECFTG  
 orf136ng AFAGTVYRFVCLFYIINDGIAHHTAPQVRVRYLFAPYRGFLPPASDSLKSSEK  
 35 orf136-1 AFVGTVYRFVCLFYIINDGIAHHSAPQVRVRYLFAPYCGFLPSASDSLKSSEK

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 67

40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 557>:

1 ATGGAAAATA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC  
 51 CGCGCCGCG TGCTTGCCG CC.TGCGGAC GCGGGGAAAT AATGTGTGCC  
 101 GCAAGCCGGT GCAAACCGCC AAACCCGCG CAGTGGTCGG TTTGGCACTC  
 151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT  
 45 201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACC TCCGCAGGTT  
 251 CGATTGTCGG CAACCTTTTT GCATCGGGTA TGTCGCCCCG CCGCCTCGAA  
 301 TTGGAAGCCG AAATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC  
 351 CACCAATGGG TTTATCAAAG GCGCAAAGCT GCAAAATTAC ATCAACCGAA  
 401 AACTCCGCG CATGCAGATT CAGCAGTTT CCATCAAATT TGCCGCC..

50 This corresponds to the amino acid sequence <SEQ ID 558; ORF137>:

1 MENMVTFSKI RPLLAIAAAA LLAAXRTAGN NAVRKPVQTA KPAAVVGLAL  
 51 GGGASKGFAH VGIKVLKEN GIPVKVVTGT SAGSIVGNLF ASGMSPDRLE  
 101 LEAEILGKTD LVDLTLSLNG FIKGAKLQNY INRKLRGMQI QQFPIKFAA..

Further work revealed the complete nucleotide sequence <SEQ ID 559>:

-323-

1 ATGGAATAA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC  
 51 CGCCGCCGCG TTGCTTGCCG CCTGCGGCAC GCGGGGAAAT AATGCTGTCC  
 101 GCAAGCCGGT GCAAACCGCC AAACCCGCCG CAGTGGTCGG TTTGGCACTC  
 151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT  
 201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT  
 251 CGATTGTCGG CAGCCTTTT GCATCGGGTA TGTCGCCCCG CCGCCTCGAA  
 301 TTGGAAGCCG AAATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC  
 351 CACCAAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA  
 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT  
 451 GCTACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AGGGGAATGC  
 501 CCGGCAGGCT GTGCGCGCTT CCGCCGCCAT TCCCAATGTG TTCCAACCCG  
 551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGTCTGTC GCAGCCCGTG  
 601 CCCGTCAAGT CCGCCCGGCG GCAGGGGGCG AATTCGTGTA TTGCCGTCGA  
 651 TATTTCCGCC CGTCCGGGCA AAAACATCAG CCAAGGTTTC TTCTCTTATC  
 701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CTGCGTTGCA AAATGAGTTG  
 751 GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGGATT TGGGTGAGT  
 801 CGGCGGATTC GATCAGAAAA AACCGGCCAT CCGGTTGGGT GAGGAGGCAG  
 851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT  
 901 TGA

20 This corresponds to the amino acid sequence <SEQ ID 560; ORF137-1>:

1 MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAVRKPVQTA KPAAVVGLAL  
 51 GGGASKGFAH VGIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSFDRLE  
 101 LEAEILGKTD LVDLTLSSTG FIKGEKLQNY INRKVGGRQI QQFPIKFAAV  
 151 ATDFETGKAV AFNQGNAQQA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV  
 201 PVSAARRQGA NFVIAVDISA RPKKNISQGF FSYLDQTLNV MSVSALQNEL  
 251 GQADVVIKPO VLDLGA VGGF DQKKRAIRLG EEAARAALPE IKRKLAAARY  
 301 \*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF137 shows 93.3% identity over a 149aa overlap with an ORF (ORF137a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf137.pep		MENMVTFSKIRPLLAIAAAAALLAAXRTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH					
35	orf137a	MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVGLALGGGASKGFAH					
		10	20	30	40	50	60
		70	80	90	100	110	120
40	orf137.pep	VGIKVLKENGIPVKVVTGTSAGSIVGNLFASGMSFDRLELEAEILGKTDLVDLTLSSTNG					
	orf137a	VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSFDRLELEAEILGKTDLVDLTLSSTG					
		70	80	90	100	110	120
45	orf137.pep	FIKGAKLQNYINRKLGRMQIQFPIKFAA					
	orf137a	FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAQAVRASAAIPNV					
		130	140	149			
		130	140	150	160	170	180

The complete length ORF137a nucleotide sequence <SEQ ID 561> is:

50 1 ATGGAATAA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC  
 51 CGCCGCCGCG TTGCTTGCCG CCTGCGGCAC GCGGGGAAAT AATGCTGTCC  
 101 GCAAGCCGGT GCAAACCGCC AAACCCGCCG CAGTGGTCGG TTTGGCACTC  
 151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT  
 201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT  
 55 251 CGATAGTCGG CAGCCTTTT GCATCGGGTA TGTCGCCCCG CCGCCTCGAA  
 301 TTGGAAGCCG AAATTTTAGG TAAAACCGAT TTGGTCGATT TAACCTTGTC  
 351 CACCAAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA  
 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT  
 451 GCTACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AAGGGAATGC  
 60 501 CGGGCAGGCT GTGCGCGCTT CCGCCGCCAT TCCCAATGTG TTCCAACCCG  
 551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGTCTGTC GCAGCCCGTG

-324-

5  
601 CCCGTCAGTG CCGCCCGGCG GCANGNNNG NATNTCGTGA TTGCCGTCGA  
651 TATTTCCGCC CGTCCGAGCA AAAACATCAG CCAAGGCTTC TTCTCTTATC  
701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CCGCGTTGCA AAATGAGTTG  
751 GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGGATT TGGGTGCAGT  
801 CGGCGGATTC GATCAGAAAA AACGCGCCAT CCGGTTGGGT GAGGAGGCAG  
851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT  
901 TGA

This encodes a protein having amino acid sequence <SEQ ID 562>:

10  
1 MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVGGLAL  
51 GGGASKGFAH VGIIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSPDRLE  
101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRKVGGRRI QQFPIKFAAV  
151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV  
201 PVSAARRXXX XXVIAVDISA RPSKNISQGF FSYLDQTLNV MSVSALQNEL  
251 QQADVVIKPO VLDLGA VGGF DQKKRAIRLG EEAARAALPE IKRKLAAARY  
15  
301 \*

ORF137a and ORF137-1 show 97.3% identity in 300 aa overlap:

20  
orf137a.pep MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVGGLALGGGASKGFAH  
orf137-1 MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAVRKPVQTAKPAAVVGGLALGGGASKGFAH  
orf137a.pep VGIIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG  
orf137-1 VGIIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG  
25  
orf137a.pep FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV  
orf137-1 FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV  
30  
orf137a.pep FQPVIIGRHTYVDGGLSQPVVSAARRXXXXXVIAVDISARPSKNISQGFFSYLDQTLNV  
orf137-1 FQPVIIGRHTYVDGGLSQPVVSAARRQGANFVIAVDISARPGKNISQGFFSYLDQTLNV  
orf137a.pep MSVSALQNELQQADVVIKPOVLDLGA VGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY  
35  
orf137-1 MSVSALQNELQQADVVIKPOVLDLGA VGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF137 shows 89.9% identity over a 149aa overlap with a predicted ORF (ORF137ng) from *N.gonorrhoeae*:

40  
orf137.pep MENMVTFSKIRPLLAIAAAAALLAAXRTAGNNAVRKPVQTAKPAAVVGGLALGGGASKGFAH 60  
orf137ng MENMVTFSKIRSFLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGGASKGFAH 60  
45  
orf137.pep VGIIKVLKENGIPVKVVTGTSAGSIVGNLFASGMSPDRLELEAEILGKTDLVDLTLSTNG 120  
orf137ng IGIVKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG 120  
orf137.pep FIKGAKLQNYINRKLGRMQIQFPIKFAA 149  
50  
orf137ng FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV 180

The complete length ORF137ng nucleotide sequence <SEQ ID 563> is:

55  
1 ATGGAATA TGGTAACGTT TTCAAAAATC AGATCATTTT TGGCAATCGC  
51 CGCCGCCGCG TTGCTTGCCG CCTGCGGTAC GGCGGGAAC AATGCCGCC  
101 GCAAGCCGGT GCAAACCGCC AAACCCGCCG CAGTGGTCGC TTTGGCACTC  
151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT ATAGGAATTG TTAAGGTTTT  
201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT  
251 CGATAGTCGG CAGCCTTTTG GCATCGGGTA TGTCGCCCGA CCGCCTCGAA  
301 TTGGAAGCCG AGATTTTAGG TAAAACCGAT TTAGTCGATT TAACCTTGTC  
351 CACCAAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA  
60  
401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT  
451 GCCACTGATT TTGAAACCGG CAAGGCCGTC GCTTCAATC AAGGGAATGC

-325-

501 CGGGCAGGCG GTTCGTGCTT CCGCCGCCAT TCCCAATGTG TTCCAGCCAG  
 551 TCATCATCGG CAGGCACAAA TATGTTGACG GCGGTCTGTC GCAGCCCGTG  
 601 CCCGTCAAGT CCGCTCGGCG GCAGGGGGCG AATTTCGTGA TTGCCGTCGA  
 651 TATTTCCGCA CGTCCGAGCA AAAATGTCGG TCAAGGTTTC TTCTCTTATC  
 701 TCGATCAGAC GCTGAACGTG ATGAGCGTTT CCGTGTGCA AAACGAGTTG  
 751 gggcAGGCGG ATGTGGTTAT CAAACCGCag gtTTTGGATT TGGGTGCAGT  
 801 CGGCGGATTC GATCAGAAAA AGCGCGCCAT CCGGTTGGGC GAGGAGGCAG  
 851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT  
 901 TGA

10 This encodes a protein having amino acid sequence <SEQ ID 564>:

1 MENMVTFSKI RSFLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVALAL  
 51 GGGASKGFAH IGIVKVLKEN GIPVKVVTGT SAGSIVGSLI ASGMSFDRLE  
 101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRKVGGRQI QFPFIKFAAV  
 151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHK YVDGGLSQPV  
 15 PVSAARRQGA NFVIAVDISA RPSKNVGQGF FSYLDQTLNV MSVSVLQNEL  
 201 GQADVVIKQ VLDLGA VGGF DQKKRAIRLG EEAARAALPE IKRKLAAYRY  
 251  
 301 \*

ORF137ng and ORF137-1 show 96.0% identity in 300 aa overlap:

20 orf137ng MENMVTFSKIRSFLAIAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGGASKGFAH  
 orf137-1 MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH  
 25 orf137ng IGIVKVLKENGIPVKVVTGTSAGSIVGSLIASGMSFDRLEAEILGKTDLVDLTLSTSG  
 orf137-1 VGIIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSFDRLEAEILGKTDLVDLTLSTSG  
 30 orf137ng FIKGEKLQNYINRKVGGRQIQFPFIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV  
 orf137-1 FIKGEKLQNYINRKVGGRQIQFPFIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV  
 35 orf137ng FQPVIIGRHKYVDGGLSQPVVSAARRQGANFVIAVDISARPSKNVGQGFSSYLDQTLNV  
 orf137-1 FQPVIIGRHTYVDGGLSQPVVSAARRQGANFVIAVDISARPGKNISQGFSSYLDQTLNV  
 40 orf137ng MSVSVLQNELGQADVVIKQVLDLGA VGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY  
 orf137 MSVSALQNELGQADVVIKQVLDLGA VGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 68

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 565>:

45 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA  
 51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCTGcTG CCGCTTTTCCT  
 101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA  
 151 AAGGAAGACC GCGCGCGCAT CGTCGCCmAT ATGCGGCAGG CGGGTTTGAA  
 201 CCCCAGCCCC AAAACGGTCA AAGCCGTTTT TGCGGAAACG GCAAAAGGCG  
 50 251 GTTTGGAAC TCCCCCGCG TTTTTCAGAA AACCGGAAGA CATAGAAACA  
 301 ATGTTCAAAG CCGTACACGG CTGGGAACAT GTGCAGCAGG CTTTGGACAA  
 351 ACACGAAGGG CTGCTATTC..

This corresponds to the amino acid sequence <SEQ ID 566; ORF138>:

1 MFRLQFRLFP PLRTAMHILL TALLKCLSLI PLSCLHTLGN RLGHlafYLL  
 51 KEDRARIVAX MRQAGLNPDF KTVKAVFAET AKGGLELAPA FFRKPEDIET  
 55 101 MFKAHVHGEH VQQALDKHEG LLF

Further work revealed the complete nucleotide sequence <SEQ ID 567>:

```

      1  ATGTTTCGTT  TACAATTCAG  GCTGTTTCCC  CCTTTGCGAA  CCGCCATGCA
     51  CATCCTGTTG  ACCGCCCTGC  TCAAATGCCT  CTCCCTGCTG  CCGCTTTCCT
    101  GTCTGCACAC  GCTGGGAAAC  CGGCTCGGAC  ATCTGGCGTT  TTACCTTTTA
    151  AAGGAAGACC  GCGCGCGCAT  CGTCGCCAAT  ATGCGGCAGG  CGGGTTTGAA
    201  CCCCAGCCCC  AAAACGGTCA  AAGCCGTTTT  TGCGGAAACG  GCAAAAGCGC
    251  GTTTGGAAC  TGCCCCGCG  TTTTTCAGAA  AACCGGAAGA  CATAGAAACA
    301  ATGTTCAAAG  CGGTACACGG  CTGGGAACAT  GTGCAGCAGG  CTTTGGACAA
    351  ACACGAAGGG  CTGCTATTCA  TCACGCCGCA  CATCGGCAGC  TACGATTGTT
    401  GCGGACGCTA  CATCAGCCAG  CAGCTTCCGT  TCCCGCTGAC  CGCCATGTAC
    451  AAACCGCCGA  AAATCAAAGC  GATAGACAAA  ATCATGCAGG  CGGGCAGGGT
    501  TCGCGGCAAA  GGAAAAACCG  CGCCTACCAG  CATACAAGGG  GTCAAACAAA
    551  TCATCAAAGC  CCTGCGTTCG  GGCGAAGCAA  CCATCGTCCT  GCCCGACCAC
    601  GTCCCTCCCT  CTCAAGAAGG  CGGGGAAGGC  GTATGGGTGG  ATTTCTTCGG
    651  CAAACCTGCC  TATACCATGA  CGCTGGCGGC  AAAATTGCGA  CACGTCAAAG
    701  GCGTGAAAC  CCTGTTTTTC  TGCTGCGAAC  GCCTGCCTGG  CGGACAAGGT
    751  TTCGATTGTC  ACATCCGCC  CGTCCAAGGG  GAATTGAACG  GCGACAAAGC
    801  CCATGATGCC  GCCGTGTTCA  ACCGCAATGC  CGAATATTGG  ATACGCCGTT
    851  TTCCGACGCA  GTATCTGTTT  ATGTACAACC  GCTACAAAAT  GCCGTAA

```

20 This corresponds to the amino acid sequence <SEQ ID 568; ORF138-1>:

```

      1  MFRLQFRLFP  PLRTAMHILL  TALLKCLSL  LPLSCLHTLGN  RLGLHAFYLL
     51  KEDRARIVAN  MRQAGLNPD  KTVKAVFAET  AKGGLELAPA  FFRKPEDIET
    101  MFKAVHGW  VQALDKHEG  LLFITPHIGS  YDLGGYISQ  QLPFPLTAMY
    151  KPPKIKAI  IMQAGRVRG  GKTAPTSIOG  VKQIIKALRS  GEATIVLPDH
    201  VSPQEGGEG  VWVDFFGKPA  YMTLAAKLA  HVKGVKTLFF  CCERLPGGQG
    251  FDLHIRPVQ  ELNGDKAHD  AVFNRAEY  W  IRRFPTQYLF  MYNRYKMP*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF138 shows 99.2% identity over a 123aa overlap with an ORF (ORF138a) from strain A of *N.*

30 *meningitidis*:

```

      10      20      30      40      50      60
orfl38.pep  MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGLHAFYLLKEDRARIVAX
      |||
orfl38a     MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGLHAFYLLKEDRARIVAN
      10      20      30      40      50      60

      70      80      90     100     110     120
orfl38.pep  MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWVQALDKHEG
      |||
orfl38a     MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWVQALDKHEG
      70      80      90     100     110     120

orfl38.pep  LLF
      |||
orfl38a     LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGTAPTSIOG
      130     140     150     160     170     180

```

The complete length ORF138a nucleotide sequence <SEQ ID 569> is:

```

      1  ATGTTTCGTT  TACAATTCAG  GCTGTTTCCC  CCTTTGCGAA  CCGCCATGCA
     51  CATCCTGTTG  ACCGCCCTGC  TCAAATGCCT  CTCCCTGCTG  CCGCTTTCCT
    101  GTCTGCACAC  GCTGGGAAAC  CGGCTCGGAC  ATCTGGCGTT  TTACCTTTTA
    151  AAGGAAGACC  GCGCGCGCAT  CGTCGCCAAT  ATGCGTCAGG  CAGGCATGAA
    201  TCCCAGCCCC  AAAACGGTCA  AAGCCGTTTT  TGCGGAAACG  GCAAAAGCGC
    251  GTTTGGAAC  TGCCCCGCG  TTTTTCAGAA  AACCGGAAGA  CATAGAAACA
    301  ATGTTCAAAG  CGGTACACGG  CTGGGAACAT  GTGCAGCAGG  CTTTGGACAA
    351  ACACGAAGGG  CTGCTATTCA  TCACGCCGCA  CATCGGCAGC  TACGATTGTT
    401  GCGGACGCTA  CATCAGCCAG  CAGCTTCCGT  TCCCGCTGAC  CGCCATGTAC
    451  AAACCGCCGA  AAATCAAAGC  GATAGACAAA  ATCATGCAGG  CGGGCAGGGT
    501  TCGCGGCAAA  GGAAAAACCG  CGCCTACCAG  CATACAAGGG  GTCAAACAAA
    551  TCATCAAAGC  CCTGCGTTCG  GGCGAAGCAA  CCATCGTCCT  GCCCGACCAC

```

5  
 601 GTCCCTCCC CTCAAGAAGG CGGGGAAGGC GTATGGGTGG ATTTCTTCGG  
 651 CAAACCTGCC TATACCATGA CGCTGGCGGC AAAATTGGCA CACGTCAAAG  
 701 GCGTGAAAAC CCTGTTTTTC TGCTGCGAAC GCCTGCCTGG CGGACAAGGT  
 751 TTCGATTTCG ACATCCGCCC CGTCCAAGGG GAATTGAACG GCGACAAAGC  
 801 CCATGATGCC GCCGTGTTCA ACCGCAATGC CGAATATTGG ATACGCCGTT  
 851 TTCCGACGCA GTATCTGTTT ATGTACAACC GCTACAAAT GCCGTAA

This encodes a protein having amino acid sequence <SEQ ID 570>:

10  
 1 MFRLQFRLFP PLRTAMHILL TALLKCLSL PLSC LHTLGN RLGH LAFYLL  
 51 KEDRARIIVAN MRQAGLNPD KTVKAVFAET AKGGL ELAPA FFRK PED IET  
 101 MFKAVHGWEH VQQA LDKHEG LLFITPHIGS YDLGG RYISQ QLPF PLTAMY  
 151 KPPKIKAIK IMQAGRVRGK GKTAPT SIQG VKQI IKA LRS GEAT IVLP DH  
 201 VPSPQEGGEG VWVDF FGKPA YTM TLA AKLA HVKG VKT LFF CCER LPGG QG  
 251 FDLHIRPVQG ELNGDKA HDA AVFN RNAEYW IRRF PTQYLF MYNRY KMP\*

ORF138a and ORF138-1 show 99.7% identity over a 298aa overlap:

15  
 orf138a.pep MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSC LHTLGNRLGH LAFYLLKEDRARIIVAN  
 orf138-1 MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSC LHTLGNRLGH LAFYLLKEDRARIIVAN  
 20  
 orf138a.pep MRQAGMNPDPKTVKAVFAETAKGGL ELAPAFFRKPEDIETMFKAVHGWEHVQQA LDKHEG  
 orf138-1 MRQAGLNPD KTVKAVFAETAKGGL ELAPAFFRKPEDIETMFKAVHGWEHVQQA LDKHEG  
 25  
 orf138a.pep LLFITPHIGSYDLGG RYISQQLPF PLTAMYKPPKIKAIKIMQAGRVRGKGKTAPT SIQG  
 orf138-1 LLFITPHIGSYDLGG RYISQQLPF PLTAMYKPPKIKAIKIMQAGRVRGKGKTAPT SIQG  
 30  
 orf138a.pep VKQI IKA LRSGEAT IVLP DHVPSPQEGGEGVWVDF FGKPAYTMT LA AKLAHVKG VKT LFF  
 orf138-1 VKQI IKA LRSGEAT IVLP DHVPSPQEGGEGVWVDF FGKPAYTMT LA AKLAHVKG VKT LFF  
 35  
 orf138a.pep CCER LPGG QGFDLHIRPVQGE LNGDKA HDA AVFN RNAEYW IRRF PTQYLF MYNRY KMP  
 orf138-1 CCER LPGG QGFDLHIRPVQGE LNGDKA HDA AVFN RNAEYW IRRF PTQYLF MYNRY KMP

### 35 Homology with a predicted ORF from *N.gonorrhoeae*

ORF138 shows 94.3% identity over a 123aa overlap with a predicted ORF (ORF138ng) from *N.gonorrhoeae*:

40  
 orf138.pep MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSC LHTLGNRLGH LAFYLLKEDRARIIVAX 60  
 orf138ng MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSC LHTLGNRLGH LAFYLLKEDRARIIVAN 60  
 45  
 orf138.pep MRQAGLNPD KTVKAVFAETAKGGL ELAPAFFRKPEDIETMFKAVHGWEHVQQA LDKHEG 120  
 orf138ng MRQAGLNPD KTVKAVFAETAKGGL ELAPAFFRKPEDIETMFKAVHGWEHVQQA LDKHEG 120  
 orf138.pep LLF 123  
 orf138ng LLFITPHIGSYDLGG RYISQQLPF PLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTGIQG 180

The complete length ORF138ng nucleotide sequence <SEQ ID 571> is:

50  
 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA  
 51 CATCCTGTTG ACCGCCCTGC TCAATATGCCT CTCCTGCTG TCGCTTTCCT  
 101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA  
 151 AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATGCGGCAGG CGGGTTTGAA  
 201 CCCCACACG CAGACGGTCA AAGCCGTTTT TGCGGAAACG GCAAAATGCG  
 55  
 251 GTTTGGA ACT TGCCCCGCG TTTTCAAAA AACCGGAAGA CATCGAAACA  
 301 ATGTTCAAAG CGGTACACGG CTGGGAACAC GTGCAGCAGG CTTTGGACAA  
 351 GGGCGAAGGG CTGCTGTTCA TCACGCCGCA CATCGGCAGC TACGATTGG  
 401 GCGGACGCTA CATCAGCCAG CAGCTTCCGT TCCACCTGAC CGCCATGTAC  
 451 AAGCCGCCGA AAATCAAAGC GATAGACAAA ATCATGCAGG CGGGCAGGGT  
 60  
 501 GCGCGGCAA GGCAAAACcg cgcaccg catACAAGGG GTCAAACAA  
 551 tcatcaAGGC CCTGCGCGCG GCGAGGCAA CCATcATCCT GCCCGACCAC

5  
 601 GTCCCTTCTC CGCAGGAagg cggCGGCGTG TGGGCGGATT TTTTCGGCAA  
 651 ACCTGCATAC acCATGACAC TGGCGGCAAA ATTGGCACAC GTCAAAGGCG  
 701 TGAAAACCTT GTTTTTCTGC TGCGAACGCC TGCCCGACGG ACAAGGCTTC  
 751 GTGTTGCACA TCCGCCCCGT CCAAGGGGAA TTGAACGGCA ACAAGCCCA  
 801 CGATGCCGCC GTGTTCAACC GCAATACCGA ATATTGGATA CGCCGTTTTC  
 851 CGACGCAGTA TCTGTTTATG TACAACCGCT ATAAAACGCC GTAA

This encodes a protein having amino acid sequence <SEQ ID 572>:

10  
 1 MFRLLQFRLFP PLRTAMHILL TALLKCLSL LSLCLHTLGN RLGHAFYLL  
 51 KEDRARIVAN MRQAGLNPD TQTVKAVFAET AKCGLELAPA FFKKPEDTET  
 101 MFKAHVHGEH VQALDKGEG LLFITPHIGS YDLGGYISQ QLPFHLTAMY  
 151 KPPKIKAIK IMQAGRVRGK GKTAPTGIQG VKQIIKALRA GEATIILPDH  
 201 VPSQEGGGV WADFFGKPAY TMTLAAKLAH VKGVKTLFFC CERLPDQGGF  
 251 VLHIRPVQGE LNGNKAHDAA VFNRNTEYWI RRFPTQYLFM YNRYKTP\*

ORF138ng and ORF138-1 show 94.3% identity over 299aa overlap:

15 orf138-1.pep MFRLLQFRLFPPLRTAMHILLTALLKCLSL LPLSCLHTLGNRLGHAFYLLKEDRARIVAN  
 orf138ng MFRLLQFRLFPPLRTAMHILLTALLKCLSL LSLCLHTLGNRLGHAFYLLKEDRARIVAN  
 20 orf138-1.pep MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDTETMFKAHVHGEHVQALDKHEG  
 orf138ng MRQAGLNPD TQTVKAVFAETAKCGLELAPAFFKKPEDTETMFKAHVHGEHVQALDKGEG  
 25 orf138-1.pep LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTISIQG  
 orf138ng LLFITPHIGSYDLGGYISQQLPFHLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTGIQG  
 30 orf138-1.pep VKQIIKALRSGEATIVLPDHVPSQEGGEGVWVDFFGKPAYTMTLAAKLAHVKGKTLFF  
 orf138ng VKQIIKALRAGEATIILPDHVPSPQEGG-GVWADFFGKPAYTMTLAAKLAHVKGKTLFF  
 35 orf138-1.pep CCERLPGGQGF DLHIRPVQGE LNDKAHDAAVFNRNAEYWI RRFPTQYLFMYNRYKMP  
 orf138ng CCERLPDQGGFVLHIRPVQGE LNDGKAHDAAVFNRNTEYWI RRFPTQYLFMYNRYKTP

In addition, ORF138ng is homologous to htrB protein from *Pseudomonas fluorescens*:

35 gnl|PID|e334283 (Y14568) htrB [*Pseudomonas fluorescens*] Length = 253  
 Score = 80.8 bits (196), Expect = 9e-15  
 Identities = 49/151 (32%), Positives = 79/151 (51%), Gaps = 6/151 (3%)  
 40 Query: 101 MFKAHVHGEHVQALDKGEGLLFITPHIGSYD-LGGYISQQLPFHLTAMYKPPKIKAIK 159  
 + + V G E + + + AL G + G + + IT H + G + + L Y S Q P Y + PPK + KA + D  
 Sbjct: 94 LVREVEGLEVLKEALASGKGVGITSHLGNWEVLNHFYCSQCKPI---IFYPKPKLKAVD 150  
 Query: 160 KIMQAGRVRGKGKTAPTGIQGVKQIIKALRAGEATIILPDHVPSPQEGGGVWADFFGKPA 219  
 + + + + RV + K A + + G + + IK + R G I D P P E G + + FF A  
 45 Sbjct: 151 ELLRKQRVQLGNKVAASTKEGILSVIKVEVRKGGQVGIPAD--PEPAESAGIFVPFFATQA 208  
 Query: 220 YMTLAAKLAHVKGKTLFFCCERLPDQGGF 250  
 T + + F RLPDG G +  
 Sbjct: 209 LTSKFVPNMLAGGKAVGVFLHALRLPDGSGY 239

50 Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF138-1 (57kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 14A  
 55 shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis



(Figure 14B). These experiments confirm that ORF138-1 is a surface-exposed protein, and that it is a useful immunogen.

### Example 69

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 573>:

```

5      1  ..GCGTGGTCGG CCGGCGAATC GTGGCGTGTG TTAATGGAAG GTGAAACGTG
      51  GCATGCGGTG TGGAACTACT TGCGCTTCTC GCGGCGGCGG GTGTATGCGG
     101  CAGCGGTTTT GGGTGTGGTG TATGCGGCGC CCGGCGGCGG GTCGCGCTGG
     151  ATGCGCGGGC TGATGTTTTA GCCGTTTATG GTGTGCGCGG TTTGTGTTTC
     201  GCGGCGCGTG CTGCTGCTTT ATCCGCACTG GACGGCTTCG TTGCCGTTGC
    110  251  TGCTGGCGAT GTATGCGCTG CTGGCGTATC CGTTTGTGGC AAAAGATGTT
     301  TTATCAGCCT GGGATGCACT GCCGCCGGAT TACGGCAGGG CCGGCGGCGG
     351  TTTGGGTGCA AACGGCTTTC AGACGGCATG CCGCATCACG TTCCCTCTCT
     401  TGAACCGGCG GTTGCGGCGC GGTCTGACTT TGGCGGCGGC AACCTGCGTG
     451  GCGCAATTTG CCGCGACATT GTTCTGTGCG CGTCCGGAAT GGCAGACGCT
    115  501  GACGACTTTG ATTTATGCCT ATTTGGGACG CGCGGGTGAG GATAATTACG
     551  CGCGGGCGAT GGTGCTG..
  
```

This corresponds to the amino acid sequence <SEQ ID 574; ORF139>:

```

      1  ..AWSAGESWRV LMESETWHAV WNTLRFSAAA VYAAAVLGTV YAAPARRSAW
     51  MRGLMFXPFM VSPVCVSAGV LLLYPQWTAS LPLLLAMYAL LAYPFVAKDV
    101  LSAWDALPPD YGRAAAGLGA NGFQTACRIT FPLLPALRR GLTLAAATCV
     151  GEFAATLFLS RPEWQTLTTL IYAYLGRAGE DNYARAMVL..
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 575>:

```

      1  ATGGATGGAC GGCCTTGGGT GGTATGGGGT GCTTTTGCCC TGCTGCCTTC
     51  GGCTTTTGTG GCGGTAATGG TCGTTGCGCC TTTGTGGGCG GTGGCGGCGT
    101  ATGACGGTTT GCGGTGGCGC GCGGTGCTGT CGGATGCCA TAATGCTCAA
    151  CGTTTGGCGT GGACGGTATT TCAGGCAGCG GCAACCTGTG TGCTGGTGCT
    201  GCCTTTGGGC GTGCTGTGCG CGTGGGTGCT GCGCGGCGTG GCGTTTCCGG
    251  GCGCGGCTTT GGTGTGCGCG CTGCTGATGC TGCCCTTTGT GATGCCCCACG
    301  TTGGTGGCGG GCGTGGGCGT GCTGGCCCTG TTCGGGGCGG ACGGGCTGTT
    351  GTGGCGCGGC AGGCAGGATA CGCGTATCT GTTGTGTGAC GGCAATGTGT
    401  TTTTCAACCT TCCTGTGTTG GTCAGGGCGG CGTATCAGGG GTTTGTGCAA
    451  GTGCCTGCGG CACGGCTTCA GACGGCACGG ACGTTGGGCG CCGGGGCGTG
    501  GCGGCGGTTT TGGGACATTG AAATGCCCGT TTTGCGCCCG TGGCTTGCCG
    551  GCGGCGGTGT CCTTGTCTTT CTGTATTGTT TTTCCGGGTT CCGGCTGGCG
    601  CTGCTGCTGG GCGGAGGCGG TTATGCCACG GTCGAAGTGG AAATTTACCA
    651  GTTGGTCATG TTCGAACTCG ATATGGCGGT TGCTTCGGTG CTGGTGTGGC
    701  TGGTGTGGG GGTAAACGGC GCGGCAGGGT TGCTGTATGC GTGGTTCGGC
    751  AGGCGCGCGG TTTCCGATAA GCGGTTTTCC CCTGTGATGC CGTCGCGCGC
    801  GCAGTCGGTC GGGGAATATG TGCTGTGCTG GTTGTGCGCG GCGGTGTGTG
    851  CTGTGTGCTG TGTTTCTCCT TTGTGTGCAA TTGTGTGTA AGCGTGTTCG
    901  GCGGCGGAAT CGTGGCGTGT GTTAATGGAA AGTGAAACGT GGCAGGCGGT
    951  GTGGAATACT TTGCGCTTCT CGGCGGCGGC GGTGTATGCG GCGGCGGTTT
   1001  TGGGTGTGGT GTATGCGGCG GCGGCGCGGC GGTGCGCGTG GATGCGCGGG
   1051  CTGATGTTTT TGCCGTTTAT GGTGTGCGCG GTTGTGTGTT CCGCGGCGGT
   1101  GCTGCTGCTT TATCCGCAGT GGACGGCTTC GTTGCCGTTG CTGCTGGCGA
   1151  TGTATGCGCT GCTGGCGTAT CCGTTGTGCG AAAAGATGT TTTATCAGCC
   1201  TGGGATGCAC TGCCGCCGGA TTACGGCAGG GCGGCGGCGG GTTTGGGTGC
   1251  AAACGGCTTT CAGACGGCAT GCCGCATCAC GTTCCCTCTC TTGAAACCGG
   1301  CGTTGCGGCG CGGTCTGACT TTGGCGGCGG CAACCTGCGT GGGCGAATTT
   1351  GCGGCGACAT GTTTCTGTGC GCGTCCGGA TGGCAGACGC TGACGACTTT
   1401  GATTTATGCC TATTTGGGAC GCGCGGGTGA GGATAATTAC GCGCGGCGGA
   1451  TGGTGTGAC ATTGTGTTG GCGGCGTTCC CGCTGGGTAT TTTCTGCTG
   1501  TTGGACGCGG GCGAAGGCGG AAAACAGACG GAAACGTTAT AA
  
```

This corresponds to the amino acid sequence <SEQ ID 576; ORF139-1>:

```

55      1  MDGRRVVWVG AFALLPSAFL AVMVVAPLWA VAAYDGLAWR AVLSDAYMLK
     51  RLAWTVFQAA ATCVLVLP LG VPVAVVLARL AFPGRALVLR LLMLPFVMP
    101  LVAGVGV LAL FGADLLWRG RQDTPYLLLY GNVFFNLVPL VRAAYQGFVQ
     151  VPAARLQTAR TLGAGAWRRF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA
  
```

-330-

201 LLLGGSRYAT VEVEIYQLVM FELDMAVASV LVWLVLGVTA AAGLLYAWFG  
 251 RRAVSDKAVS PVMPSPPQSV GEYVLLAFAA AVLSVCCLFP LLAIIVKAWs  
 301 AGESWRVLME SETWQAVWNT LRFSAAAVYA AAVLGVVYAA AARRSAWMRG  
 351 LMFLPFMVSP VCVSAGVLLL YPQWTASLPL LLAMYALLAY PFVAKDVLSA  
 401 WDALPPDYGR AAAGLGANGF QTACRITFPL LKPALRRGLT LAAATCVGEF  
 451 AATLFLSRPE WQTLTTLIYA YLGRAGEDNY ARAMVLTLLL AAFALGIFLL  
 501 LDGGEGGKQT ETL\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

10 ORF139 shows 94.7% identity over a 189aa overlap with an ORF (ORF139a) from strain A of *N. meningitidis*:

					10	20	30
	orf139.pep				AWSAGESWRVLMESETWHA	VWNTLRFSAAA	
15	orf139a	QSVGEYVLLAFAAAVXSVCCLFXLLAIIVK	AWsAGESWRVLMESETWQAVWNTXRFSA				
		270	280	290	300	310	320
	orf139.pep						
20	orf139a	VYAAAVLGVVYAAAPARRSAWMRGLMFXPFMVSPVCVSAGVLLLYPQWTASLPLLLAMYAL					
		330	340	350	360	370	380
	orf139.pep						
25	orf139a	LAYPFVAKDVLSAWDALPPDYGRAAAGLGANGFQTACRITFPLLPALRRGLTLAAATCV					
		390	400	410	420	430	440
	orf139.pep						
30	orf139a	GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL					
		450	460	470	480	490	500
	orf139a	GEFAATLFXSRXEWQTLTTLIYAYXGRAGXDNYARAMVLTLLLAFAALGXFLLLDGGEGG					

35 The complete length ORF139a nucleotide sequence <SEQ ID 577> is:

	1	ATGGATGGAC	GGCGTTGGGC	GGTATGGGGT	GCTTTTGCCC	TGCTGCCTTC
	51	GGCTTTTTTG	GCGGCAATGG	TCGTTGCGCC	TTGTGGGCG	GTGGCGGCGT
	101	ATGACGGTTT	GGCGTGGCGC	GCGTGCTGT	CGGATGCCTA	TATGCTCAAA
	151	CGTTTGGCGT	GGACGGTATT	TCAGGCAGCG	GCAACCTGTG	TGCTGGTGCT
40	201	GCCTTTGGGC	GTGCCTGTCG	CGTGGTGCT	GGCGCGGCTG	GCGTTCCGG
	251	GGCGGGCTTT	GGTGTGCGC	CTGTGATGC	TGCCTTTTGT	GATGCCACG
	301	TTGGTGGCGG	GCGTGGGCGT	GCTGGCTCTG	TTGGGGGCGG	ACGGCCTGTN
	351	GTGGCGCGGC	TGGCAGGATA	CGCCGTATCT	GTTGTTGTAC	GGCAATGTGT
	401	TTTTTNACCT	TCCTGTGTTG	GTCAGGGCGG	CATATCAGGG	GTTGTGCAA
45	451	GTGCCTGCGG	CACGGCTTCA	GACGGCACNG	ACATTGGGCG	CGGGGGCGTG
	501	GCGGCGGTTT	TGGGACATTG	AAATGCCCGT	TTTGGCGCCG	TGGCTTGCCG
	551	GCGGCGGTG	CCTTGTCTTC	CTGTATGTT	TTTGGGGGTT	CGGGCTGGCA
	601	TTGCTGCTGG	GCGGCAGCCG	TTATGCCACG	GTCGAAGTGG	AAATTTACCA
	651	GTTGGTCATG	TTCGAAGTCG	ATATGGCGGT	TGCTTCGGTG	CTNGTGTGGC
50	701	TGGTGTNGGG	GGTAACNGCG	GCGGCAGGGT	TGCTGTATGC	GTGGTTCGGC
	751	AGGCGCGCGG	TTTCGGATAA	GGCNGTTTCC	CCTGTGATGC	CGTCGCCGCC
	801	GCAGTCGGTC	GGGGAATATG	TGCTNCTGGC	GTTTGGCGCG	GCGGTGTNGT
	851	CTGTGTGCTG	CCTGTTTCNT	TTGTTGGCAA	TTGTTGTGAA	AGCGTGGTCG
	901	CGCGGCGAAT	CGTGGCGTGT	GTTAATGGAA	AGTGAACGTT	GGCAGGCGGT
55	951	GTGGAATACT	NTGCGCTTCT	CGGCGGCGGC	GGTGTATGCG	GCGGCGGTTT
	1001	TGGGTGTGGT	GTATGCGCGG	GCGGCGCGGC	GGTGGGCGTG	GATGCGCGGG
	1051	CTGATGTTT	TGCCGTTTAT	GGTGTGCGCG	GTTGTGTGTT	CGGCGGGCGT
	1101	GCTGTGCTT	NATCCGCAGT	GGACGGCTTC	GTTGCCGCTG	CTGCTGGCGA
	1151	TGTATGCGCT	GCTGGCGTAT	CCGTTGTGTT	CAAAAGATGT	TTTATCAGCC
60	1201	TGNGATGCAC	TGCCGCCGGA	TTACGGCAGG	GCGGCGGCGG	GTTTGGGTGC
	1251	AAACGGCTTT	CAGACGGCAT	GCCGCATCAC	GTTCCCCCTC	TTGAAACCGG
	1301	CGTTGCGGCG	CGGTCTGACT	TTGGCGGCGG	CAACCTGCGT	GGGCGAATTT
	1351	GCGGCAACCT	TGTTCTNTGTC	GCGTCNCGAG	TGGCAGACGC	TGACGACTTT

5	1	MDGRRWAVWG	AFALLPSAFL	AAMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFQAA	ATCVLVPLG	VPVAWVLARL	AFPGRALVLR	LLMLPFVMP
	101	LVAGVGVLLAL	FGADGLXWRG	WQDTPYLLLY	GNVFXLPVL	VRAAYQGQFVQ
	151	VPAARLQTXA	TLGAGAWRRF	WDIEMPVLRP	WLAGGVCLVF	LYCFSGGFGLA
10	201	LLLGSRVYAT	VEVEIYQLVM	FELDMAVASV	LVVLVXGVTA	AAGLYLAWFG
	251	RRAVSDKAVS	PVMPSPQSV	GEYVLLAFAA	AVXSVCCLEF	LLAIVVKAWS
	301	AGESWRVLME	SETWQAVWNT	XRFSAAVYA	AAVLGVVYAA	AAARSAWMRG
	351	LMFLPFMVSP	VCVSAGVLLL	XPQWTASLPL	LLAMYALLAY	PFVAKDVLSA
15	401	XDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAATCVGEF
	451	AATLFXSRXE	WQTLTTLIYA	YXGRAGXDNY	ARAMVLTLLL	AAFALGXFLL
	501	LDGGEGGKRT	ETL*			

	orf139a.pep	MDGRRWAVWGAFALLPSAFLAAMVVAAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA      :     :
20	orf139-1	MDGRRWVVGAFALLPSAFLAVMVVAAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
	orf139a.pep	ATCVLVLPPLGVPVAWVLARLAFPGRALVLRLLMLPFVMPPTLVAGVGVLALFGADGLXWRG 
	orf139-1	ATCVLVLPPLGVPVAWVLARLAFPGRALVLRLLMLPFVMPPTLVAGVGVLALFGADGLLWRG
25	orf139a.pep	WQDTPYLLLYGNVFFXLPVLVRAAYQGFGVQVPAARLQTAXTLGAGAWRRFWDIEMPVLRP 
	orf139-1	RQDTPYLLLYGNVFFNLVPLVRAAYQGFGVQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
30	orf139a.pep	WLAGGVCLVFLYCFSGFGLALLLGGSSRYATVEVEIYQLVMFELDMAVASVLVWLXGVTA 
	orf139-1	WLAGGVCLVFLYCFSGFGLALLLGGSSRYATVEVEIYQLVMFELDMAVASVLVWLVLGVTA
	orf139a.pep	AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFAAAVXSVCCLFXLLAIVVKAWS 
35	orf139-1	AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFAAAVLSVCCLFPLLAIVVKAWS
	orf139a.pep	AGESWRVLMESETWQAVWNTXRFSAAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP 
40	orf139-1	AGESWRVLMESETWQAVWNTLRFSAAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP
	orf139a.pep	VCVSAGVLLLPQWTASLPLLLAMYALLAYPFVAKDVL SAXDALPPDYGRAAAGLGANGF 
	orf139-1	VCVSAGVLLLPQWTASLPLLLAMYALLAYPFVAKDVL SAWDALPPDYGRAAAGLGANGF
45	orf139a.pep	QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFXSRXEWQTLTTLIYAYXGRAGXDNY 
	orf139-1	QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
	orf139a.pep	ARAMVLTLLLAFAALGXFLLLDGGEGGKRTETLX 
50	orf139-1	ARAMVLTLLLAFAALGIFLLDGGEGGKQTETLX

55 *N.gonorrhoeae*:

	orf139.pep	AWSAGESWRVLMSESETWHAVWNTLRFSA	30
		:	
	orf139ng	QSVGEYVLLAFSAVLVSVCLFPLSAIVVKAWSAGESRRVLMSESETWQAVWNTLRFSA	327
60	orf139.pep	VYAAAVLGVVYAAPARRSAWMRGLMFXPFMVSPVCVSAGVLLLYPQWTASLPLLLAMYAL	90
		:       :	
	orf139ng	VFAAAVLGVVYAAAARLVWMRGLVFLPFMVSPVCVSAGVLLLYPGWTASLPLLLAMYAL	387

orf139.pep	LAYPFVAKDVLSAWDALPPDYGRAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV	150
orf139ng	LAYPFVAKDVLSAWDALPPDYGRAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV	447
5 orf139.pep	GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL	189
orf139ng	GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVLTLLLSAFAVCIFLLLDNDEGG	507

The complete length ORF139ng nucleotide sequence <SEQ ID 579> is predicted to encode a protein having amino acid sequence <SEQ ID 580>:

10	1	MDGRCWAVRG	AFSLLEPSAFL	AVMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFQAA	ATCVLVLPLG	VPVAWVLARL	AFPGRALVLR	LLMLPFVMP
	101	LVAGVGVLAL	FGADGLLWRG	RQDTPYLLLY	GNVFFNLPLV	VRAAYQGFAQ
	151	VPAARLQTAR	TLGAGAWRPF	WDIEMPVLRP	WLAGGVCLVF	LYCFSGFGLA
	201	LLLGGSTRYAT	VEVEIYQLVM	FELDMAGASA	LVWLVLGVTA	AAGLLYAWFG
15	251	RRAVSDKAVS	PVMPSPQSV	GEYVLLAFSV	AVLSVCCLFP	LSAIVVKAWS
	301	AGESRRVLME	SETWQAVWNT	LRFSAAAVFA	AAVLGVVYAA	AARRLVWMRG
	351	LVFLPFMVSP	VCVSAGVLLL	YPGWTASLPL	LLAMYALLAY	PFVAKDVLSA
	401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAAATCVGEF
	451	AATLFLSRPE	WQTLTTLIYA	YLGRAGEDNY	ARAMVLTLLL	SAFAVCIFLL
20	501	LDNDEGGKRT	ETL*			

Further work revealed a variant gonococcal DNA sequence <SEQ ID 581>:

	1	ATGGATGGAC	GGTGTGGGCG	GGTACGGGGT	GCTTTTCCCG	TGCTGCCTTC
	51	GGCTTTTGTG	GCGGTAATGG	TCGTTGCGCC	TTTGTGGGCG	GTGGCGGCGT
	101	ATGACGGTTT	GCGGTGGCGC	GCGGTGCTGT	CGGATGCCTA	TATGCTCAAA
25	151	CGTTTGGCGT	GGACGGTGT	TCAGGCGGCG	GCAACCTGTG	TGCTGGTGCT
	201	GCCTTTGGGCG	GTGCCTGTG	CGTGGGTGCT	GGCGCGGCTG	GCGTTCCTCG
	251	GGCGGGCTTT	GGTGTGCGC	CTGCTGATGC	TGCCGTTTGT	GATGCCACG
	301	CTGTGCGCG	GCGTGGGCTG	GCTGGCTCTG	TTCGGGGCGG	ACGGGTGTTT
	351	GTGGCGCGG	CGGCAGGATA	CGCCGTATCT	GTTGTTGTAC	GGCAATGTGT
30	401	TTTTCAACCT	GCCCGTGTG	GTCAGGCGG	CGTATCAGG	GTTTGCTCAA
	451	GTGCCTGCGG	CACGGCTTCA	GACGGCACGG	ACGTTGGGCG	CGGGGGCGTG
	501	GCGGCGGTTT	TGGGACATTG	AAATGCCCCG	TTTGCGCCCC	TGGCTTGCCG
	551	GCGGCGGTG	CCTGTCTTTC	CTGTATTGTT	TTTCGGGGTT	CGGGGTGGCA
	601	TTGCTGTTGG	GCGGCAGCCG	TTATGCCACG	GTCGAAGTGG	AAATTTACCA
35	651	GTTGTTATG	TTCGAACCTG	ATATGGCGGG	GGCTTCGGCG	CTGGTGTGGC
	701	TGGTGTGGG	GGTAACGGCG	GCGGCAGGGT	TGCTGTATGC	GTGGTTCGCG
	751	AGGCGCGCG	TTTCGGATAA	GGCGGTTTCC	CCCGTGATGC	CGTCGCCGCC
	801	GCAATCGGTG	GGGGAATATG	TATTGCTGGC	ATTTTCGGTG	GCGGTGTTGT
	851	CCGTGTGCTG	CCTGTTTCCT	TTGTGCGCAA	TTGTTGTGAA	AGCGTGGTCG
40	901	GCCGCGCAAT	CGCGCGGTG	GTTAATGGAA	AGTGAAACGT	GGCAGGCAGT
	951	GTGAATACT	ttGCGCTTTT	GCGCGCGCGC	GGTGTTCGCG	GCGGCGGTTT
	1001	TGGGTGTGGT	GTATGCGGCG	GCGGCGCGGC	GGCTGGTGTG	GATGCGCGGA
	1051	CTGTGTGTTT	TACGTTTAT	GGTGTGCGCG	GTTTGTGTTT	GCGGCGGCGT
	1101	GCTGCTGCTT	TATCCGGGGT	GGACGGCTTC	GTTACCGCTG	CTGCTGGCGA
45	1151	TGTATGCGCT	GCTGCGGTAT	CCGTTTGTGG	CAAAAGATGT	TTTATCGGCC
	1201	TGGGATGCAC	TGCCGCCGGA	TTACGGCAGG	GCGGCGGCAG	GTTTGGGCGC
	1251	AAACGGCTTT	CAGACGGCAT	GCCGTATCAC	GTTCCCCCTC	TTGAAACCGG
	1301	CGTTGCGGCG	CGGTCTGACT	TTGGCGGCGG	CGACGTGTGT	GGGCGAATTT
	1351	GCGGCAACCT	TGTTCTGTG	GCGTCCGGAA	TGGCAGACGT	TGACGACTTT
50	1401	GATTTATGCC	TATTTGGGGC	GTGCGGGTGA	GGACAATTAT	GCGGCGGCAA
	1451	TGGTGTGAC	ATTGCTGTTG	TCGGCATTGT	CGGTGTGCAT	TTTCCTGCTG
	1501	TTGGACAACG	GCGAAGGCGG	aaaACGGACG	GAAACGTTAT	AA

This corresponds to the amino acid sequence <SEQ ID 582; ORF139ng-1>:

55	1	MDGRCWAVRG	AFSLLEPSAFL	AVMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFQAA	ATCVLVLPLG	VPVAWVLARL	AFPGRALVLR	LLMLPFVMP
	101	LVAGVGVLAL	FGADGLLWRG	RQDTPYLLLY	GNVFFNLPLV	VRAAYQGFAQ
	151	VPAARLQTAR	TLGAGAWRRF	WDIEMPVLRP	WLAGGVCLVF	LYCFSGFGLA
	201	LLLGGSTRYAT	VEVEIYQLVM	FELDMAGASA	LVWLVLGVTA	AAGLLYAWFG
	251	RRAVSDKAVS	PVMPSPQSV	GEYVLLAFSV	AVLSVCCLFP	LSAIVVKAWS
60	301	AGESRRVLME	SETWQAVWNT	LRFSAAAVFA	AAVLGVVYAA	AARRLVWMRG
	351	LVFLPFMVSP	VCVSAGVLLL	YPGWTASLPL	LLAMYALLAY	PFVAKDVLSA
	401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAAATCVGEF
	451	AATLFLSRPE	WQTLTTLIYA	YLGRAGEDNY	ARAMVLTLLL	SAFAVCIFLL
	501	LDNDEGGKRT	ETL*			

ORF139ng-1 and ORF139-1 show 95.9% identity over 513aa overlap:

```

5  orf139ng      MDGRCWAVRGAFSLPSAFLAVMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
   orf139-1      MDGRRWVWVGAFALLPSAFLAVMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
10 orf139ng      ATCVLVLPPLGVPAWVLARLAFPGRALVRLMLPFVMPTLVAGVGVLALEFGADGLLWRG
   orf139-1      ATCVLVLPPLGVPAWVLARLAFPGRALVRLMLPFVMPTLVAGVGVLALEFGADGLLWRG
15 orf139ng      RQDTPYLLLYGNVFFNLPVLVRAAYQGFAQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
   orf139-1      RQDTPYLLLYGNVFFNLPVLVRAAYQGFVQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
20 orf139ng      WLAGGVCLVFLYCFSGFGLALLGGSRYATVEVEIYQLVMFELDMAGASALVWLVLGVTA
   orf139-1      WLAGGVCLVFLYCFSGFGLALLGGSRYATVEVEIYQLVMFELDMASVWLVLVLGVTA
25 orf139ng      AAGLLYAWFGRRAVSDKAVSPVMPSPQSVGEYVLLAFSAVLSVCCLFPLSAIVVKAWS
   orf139-1      AAGLLYAWFGRRAVSDKAVSPVMPSPQSVGEYVLLAFSAVLSVCCLFPLSAIVVKAWS
30 orf139ng      AGESRRVLMESETWQAVWNTLRFSAAAVFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSP
   orf139-1      AGESRRVLMESETWQAVWNTLRFSAAAVFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSP
35 orf139ng      VCVSAGVLLLYPGWTASLPLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAGLGANGF
   orf139-1      VCVSAGVLLLYPGWTASLPLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAGLGANGF
40 orf139ng      QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
   orf139-1      QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
45 orf139ng      ARAMVLTLLLSAFAVCIFLLLDNGEGGKRTETL
   orf139-1      ARAMVLTLLLSAFAVCIFLLLDNGEGGKRTETL

```

Based on the presence of a predicted binding-protein-dependent transport systems inner membrane component signature (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 70

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 583>:

```

45 1 ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTC
   51 GCGGCGCGCA ATCATCCTCA TTCTGATTTT AATCGTCAGA TTCCGCATCC
   101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC
   151 GGTTCGCCCA CAGGCAGCAT TGTCAAAGAC ATACTGGTCA AAAACTTCGG
   201 CGGCACGCTC GCGGCGGTGG CGCTTCTGGT CCGCCTGGGC GCGATGCTCG
   251 AACGTTTGGT C...

```

This corresponds to the amino acid sequence <SEQ ID 584; ORF140>:

```

50 1 MDGWTQTLSA QTLIGISAAA IILILILIVR FRIHALLTLV IVSLLTALAT
   51 GLPTGSIVKD ILVKNFGGTL GGVALLVGLG AMLERLV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 585>:

```

55 1 ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTC
   51 GCGGCGCGCA ATCATCCTCA TTCTGATTTT AATCGTCAA TTCCGCATCC
   101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC
   151 GGTTCGCCCA CAGGCAGCAT TGTCAAAGAC ATACTGGTCA AAAACTTCGG

```

-334-

201 CGGCACGCTC GCGGCGGTGG CGCTTCTGGT CCGCCTGGGC GCGATGCTCG  
 251 GACGTTTGGT CGAAACATCC GCGGCGGCAC AGTCGCTGGC GGACGCGCTG  
 301 ATCCGGATGT TCGCGCAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCTC  
 351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC  
 5 401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAAACAGGA CGTACTGCCC  
 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCTGCCC  
 501 GCCCCATCCG GGGCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG  
 551 GCCAAGTTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTC  
 601 AGCGGCTATA TGCTCGGCAA AGTGTGCGG GCGACCATCC ATGTTCCCGT  
 10 651 TCCCGAAGTG CTCAGCGGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC  
 701 CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATTCC CATGCTGCTG  
 751 ATTTTCTCTGA ATACCGGCGT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG  
 801 TGCGGACGAA ACCTGGGTTC AGACGGCAAA AATAATCGGT TCGACACCGA  
 851 TCGCCCTTCT GATTTCCGTA TTGGTTCGAC TGTTCGTCTT GGGACGCAAA  
 15 901 CCGCGCGAAA GCGGCAGCGC GTTGGAAAA ACCGTGGACG GCGCACTCGC  
 951 CCCGTCTGT TCCGTGATTG TGATTACCGG CGCGGGCGGT ATGTTCCGCG  
 1001 GCGTTTTGCG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG  
 1051 GATTTGGGCA TTCCCGTCTT TTTGGGCTGT TTCCTTGTCT CTTGGCACT  
 1101 GCGTATCGCG CAAGGTTCGG CAACCGTCGC CCTGACCACC GCGCCGCGC  
 20 1151 TGATGGCTCC TGCCGTTGCC GCGCGCGGCT TTACCGACTG GCAGCTCGCC  
 1201 SGTATCGTAT TGCAACGGC GGCAGGTTTC GTCGGTTGCA GCCACTTCAA  
 1251 CGACTCCGGC TTCTGGCTGG TCGGCCGTCT CTTGGACATG GACGTACCGA  
 1301 CCACGCTGAA AACCTGGACG GTCAACCAAA CCCTCATCGC ACTCATCGGC  
 1351 TTGCCTTGT CCGCACTGCT GTTCGCCATC GTCTGA

25 This corresponds to the amino acid sequence <SEQ ID 586; ORF140-1>:

1 MDGWTQTLA QTLGISAAA IILILILIVK FRIHALTLV IVSLLTALAT  
 51 GLPTGSIVND ILVKNFGGTL GGVALLVGL AMLGRIVETS GGAQSLADAL  
 101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP  
 151 FALASIGAFS VMHVLPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF  
 30 201 SGYMLGKVLG RTHVPVPEL LSGGTQDNLD PKEPAKAGTV VAIMLIPMLL  
 251 IFLNTGVSALE ISEKLVSAD TWWQTAKIIG STPIALLISV LVALFVLGRK  
 301 RGESESALEK TVDGAALPVC SVILITGAGG MFGGVLRASG IGKALADSMA  
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA  
 401 CIVLATAAGS VGCSEFNDSG FWLVGRLLDM DVPTTLKWT VNQTLLIALIG  
 35 451 FALSALLFAI V\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF140 shows 95.4% identity over a 87aa overlap with an ORF (ORF140a) from strain A of *N. meningitidis*:

40	orfl40.pep	MDGWTQTLAQTLLGISAAAIIILILILIVRFRIHALTLVIVSLLTALATGLPTGSIVKD
	orfl40a	MDGWTQTLAQTLLGISAAAIIILILILIVKFRHALTLVIVSLLTALATGLPTGSIVND
45	orfl40.pep	ILVKNFGGTLGGVALLVGLGAMLERLV
	orfl40a	VLVKNFGGTLGGVALLVGLGAMLRVETSSGGAQSLADALIRMFGEKRAPFALGVASLIF

The complete length ORF140a nucleotide sequence <SEQ ID 587> is:

1 ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTC  
 51 GCGGCGGGCA ATCATCCTCA TTCTGATTTT AATCGTCAAA TTCCGCATCC  
 101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC  
 55 151 GGTTCGCCCA CAGGCAGCAT TGTCACAGAC GACTGGTCA AAAACTTCGG  
 201 CCGCACGCTC GCGGCGGTGG CGCTTCTGGT CCGCCTGGGC GCGATGCTCG  
 251 GACGTTTGGT CGAAACATCC GCGGCGGCAC AGTCGCTGGC GGACGCGCTG  
 301 ATCCGGATGT TCGCGCAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCTC  
 351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC  
 60 401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAAACAGGA CGTACTGCCC  
 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCTGCCC

-335-

501 GCCCCATCCG GGCCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG  
 551 GCCAAGTTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTT  
 601 AGCGGCTATA TGCTCGGCAA AGTGTGGGG CGCACCATCC ATGTTCCCGT  
 5 651 TCCCGAAGTG CTCAGCGGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC  
 701 CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATTCC CATGCTGCTG  
 751 ATTTTCTGA ATACCGGCGT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG  
 801 TCGCGACGAA ACCTGGGTTC AGACGGCAAA AATAATCGGT TCGACACCGA  
 851 TCGCCCTTCT GATTTCCGTA TTGGTCGCAC TGTTTGTCTT GGGACGCAAA  
 901 CGCGGCGAAA GCGGCAGCGC GTTGGAAAAA ACCGTGGACG GCGCACTCGC  
 10 951 CCCGTCTGT TCCGTGATTG TGATTACCGG CGCGGGCGGT ATGTTGGCGG  
 1001 GCGTTTTGCG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG  
 1051 GATTTGGGCA TTCCCGTCCT TTTGGGCTGT TTCCTTGTCG CCTTGGCACT  
 1101 GCGTATCGCG CAAGGTTCGG CAACCGTCGC CCTGACCACC GCCGCCGCGC  
 1151 TGATGGCTCC TGCCGTGGCC GCCGCCGGCT TTACCGACTG GCAGCTCGCC  
 15 1201 TGTATCGTAT TTGCAACGGC GGCAGGTCG GTCGGTTGCA GCCACTTCAA  
 1251 CGACTCCGGC TTCTGGCTGG TCGGCCGCT CTTGGACATG GACGTACCGA  
 1301 CCACGCTGAA AACCTGGACG GTCAACCAAA CCCTCATCGC ACTCATCGGC  
 1351 TTTGCCTTGT CCGCACTGCT GTTCGCCATC GTCTGA

This encodes a protein having amino acid sequence <SEQ ID 588>:

20 1 MDGWTQTLA QTLGISAAA IILILILIVK FRIHALLTLV IVSLLTALAT  
 51 GLPTGSIVND VLVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL  
 101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP  
 151 FALASIGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF  
 201 SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV VAIMLIPMLL  
 25 251 IFLNTGVSAL ISEKLVSADE TWVQTAKIIG STPIALLISV LVALFVLGRK  
 301 RGECSALEK TVDGALAPVC SVILITGAGG MFGGVLRASG IGKALADSMA  
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA  
 401 CIVLATAAGS VGCSEFNDSD FVLVGRLLDM DVPTTLKTWT VNQTLIALIG  
 451 FALSALLFAI V\*

30 ORF140a and ORF140-1 show 99.8% identity over a 461aa overlap:

orf140-1.pep MDGWTQTLAQTLLGISAAAIIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60  
 orf140a MDGWTQTLAQTLLGISAAAIIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60  
 35 orf140-1.pep ILVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF 120  
 orf140a VLVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF 120  
 40 orf140-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASIGAFSVMHVFLPPHPGPIAASEFYG 180  
 orf140a GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASIGAFSVMHVFLPPHPGPIAASEFYG 810  
 orf140-1.pep ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTHVPVPELLSGGTQDNDLPKEPAKAGTV 240  
 45 orf140a ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTHVPVPELLSGGTQDNDLPKEPAKAGTV 240  
 orf140-1.pep VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVALFVLGRK 300  
 orf140a VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVALFVLGRK 300  
 50 orf140-1.pep RGECSALEKTVDGALAPVCSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC 360  
 orf140a RGECSALEKTVDGALAPVCSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC 360  
 55 orf140-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSEFNDSD 420  
 orf140a FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSEFNDSD 420  
 60 orf140-1.pep FWLVGRLLDMDVPTTLKTWTVNQTLIALIGFALSALLFAIV 461  
 orf140a FWLVGRLLDMDVPTTLKTWTVNQTLIALIGFALSALLFAIV 461

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF140 shows 92% identity over a 87aa overlap with a predicted ORF (ORF140ng) from

65 *N.gonorrhoeae*:

```

orfl140.pep  MDGWTQTLQAQTLGISAIAIILILILIVFRIRALLTLVIVSLLTALATGLPTGSIVKD  60
              ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orfl140ng    MDGRTQTLQAQTLGISAIAIILILILIVKFRIRALLTLVIASLLTALATGLPTGSIVND  60

5    orfl140.pep  ILVKNFGGTLGGVALLVGLGAMLERLV  87
              :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orfl140ng    VLVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFAPGVASLIF  120

```

The complete length ORF140ng nucleotide sequence <SEQ ID 589> was predicted to encode a protein having amino acid sequence <SEQ ID 590>:

```

10      1  MDGRTQTLQA  QTLGISAIA  IILILILIVK  FRIRALLTLV  IASLLTALAT
          51  GLPTGSIVND  VLVKNFGGTL  GGVALLVGLG  AMLGRLVETS  GGAQSLADAL
        101  IRMFGEKRAP  FAPGVASLIF  GFPIFFDAGL  IVMLPIVFAT  ARRMKQDVLP
        151  FALASVGAFA  VMHVFLPPHP  GPIAASEFYG  ANIGQVLILG  LPTAFITWYF
        201  SGYMLGKVLG  RAIHVPVPEL  LSGGTQDSDP  PKEPAKAGTV  VAVMLIPMLL
15      251  IFLNTGVSAL  ISEKLVSAD  TWVQTAKMIG  STPVALLISV  LAALLVLGRK
        301  RGESEGSTLEK  TVDGALAPAC  SVILITGAGG  MFGGVLRASG  IGKALADSMA
        351  DLGIPVLLGC  FLVALALRIA  QGSATVALTT  AAALMAPAVA  AAGFTDWQLA
        401  CIVLATAAGS  VGCSEFND  FWLVGRLLSDM  DVPTTLKTWT  VNQTLLIAFIG
        451  FALSALLFAI  V*

```

20 Further work revealed a variant gonococcal DNA sequence <SEQ ID 591>:

```

          1  ATGGACGGCC  GGACACAGAC  GCTGTCCGCG  CAAACCTTGT  TGGGCATTTC
          51  GCGCGCGGCA  ATCATCCTCA  TTCTGATTTT  AATCGTCAAA  TTCCGCATCC
        101  GCGCGCTGCT  GACACTGGTC  ATCGCCAGCC  TGCTGACGGC  TTTGGCAACC
        151  GGTTCGCCCA  CAGCAGCAT  CGTCAACGAC  GTACTGGTCA  AAAACTTCGG
25      201  CGGCACGCTC  GCGCGCGTGG  CGCTTCTGGT  CGGTCTGGGC  GCAATGCTCG
        251  GACGTTTGGT  AGAAACATCC  GCGCGCGCAC  AGTCGCTGCG  GGACGCGCTG
        301  ATCCGGATGT  TCGCGCAAAA  ACGCGCACCG  TTCGCTCCGG  GCGTTGCCCTC
        351  GCTGATTTTC  GGCTTCCCGA  TTTTCTTCGA  TGCCGGACTA  ATCGTCATGC
        401  TGCCCATCGT  ATTCGCCACC  GCACGGCGCA  TGAAACAGGA  CGTACTGCCC
30      451  TTCGCGCTTG  CCTCCGTCGG  CGCATTTTCC  GTCATGCACG  TCTTCCTGCC
        501  GCCCCATCCG  GGCCCGATTG  CCGCTTCCGA  ATTTTACGGC  GCGAACATCG
        551  GCCAGGTTTT  GATTTTGGGT  CTGCCGACCG  CCTTCATCAC  ATGGTATTTT
        601  AGCGCTATA  TGCTCCGCAA  AGTGTTGGGG  CGCGCCATCC  ATGTTCCCGT
        651  TCCGAACTG  CTCAGCGGCG  GCACGCAAGA  CAGCGACCCG  CCGAAAGAAC
35      701  CTGCCAAAGC  AGGAACGGTC  GTCGCCGTCA  TGCTGATTCC  CATGCTGCTG
        751  ATTTTCCTGA  ATACCGGCGT  ATCAGCCCTC  ATCAGCGAAA  AACTCGTAAG
        801  TCGCGACGAA  ACTTGGGTTT  AGACGGCAAA  AATGATCGGT  TCGACACCTG
        851  TCGCCCTTCT  GATTTCCGTA  TTGGCCGCAC  TGTTGGTCTT  GGGACGCAAA
        901  CGCGCGGAAA  GCGGCAGCAC  GTTGAAAAA  ACCGTGGACG  GCGCACTCGC
40      951  CCCCCTGT  TCCGTGATT  TGATTACCG  CGCGGGCGGT  ATGTTCCGGC
        1001  GCGTTTTCG  CGCTTCCGGC  ATCGGCAAGG  CACTCGCCGA  CAGCATGCGC
        1051  GATTTGGGCA  TTCCCGTCT  TTTGGGCTGC  TTCCTTGTCG  CCTTGGCACT
        1101  GCGTATCGCG  CAAGGTTCCG  CAACCGTCG  CCTGACCACA  GCCGCCGCGC
        1151  TGATGGCTCC  TGCCGTGTC  GCCGCCGGCT  TTACCGACTG  GCAGCTCGCC
45      1201  TGTATCGTAT  TGGCAACGGC  GGCAGGTTTC  GTCGGTTGCA  GCCACTTCAA
        1251  CGACTCCGGC  TTCTGGCTGG  TCGGCCGCCT  CTTGGATATG  GACGTACCGA
        1301  CCACGCTGAA  AACCTGGACG  GTCAACCAA  CCCTCATCGC  ATTCATCGGC
        1351  TTTGCCTTGT  CCGCACTGCT  GTTGCCATC  GTCTGA

```

This corresponds to the amino acid sequence <SEQ ID 592; ORF140ng-1>:

```

50      1  MDGRTQTLQA  QTLGISAIA  IILILILIVK  FRIRALLTLV  IASLLTALAT
          51  GLPTGSIVND  VLVKNFGGTL  GGVALLVGLG  AMLGRLVETS  GGAQSLADAL
        101  IRMFGEKRAP  FAPGVASLIF  GFPIFFDAGL  IVMLPIVFAT  ARRMKQDVLP
        151  FALASVGAFA  VMHVFLPPHP  GPIAASEFYG  ANIGQVLILG  LPTAFITWYF
        201  SGYMLGKVLG  RAIHVPVPEL  LSGGTQDSDP  PKEPAKAGTV  VAVMLIPMLL
55      251  IFLNTGVSAL  ISEKLVSAD  TWVQTAKMIG  STPVALLISV  LAALLVLGRK
        301  RGESEGSTLEK  TVDGALAPAC  SVILITGAGG  MFGGVLRASG  IGKALADSMA
        351  DLGIPVLLGC  FLVALALRIA  QGSATVALTT  AAALMAPAVA  AAGFTDWQLA
        401  CIVLATAAGS  VGCSEFND  FWLVGRLLSDM  DVPTTLKTWT  VNQTLLIAFIG
        451  FALSALLFAI  V*

```

60 ORF140ng-1 and ORF140-1 show 96.3% identity over 461aa overlap:

```

orfl140ng-1.pep MDGRTQTLQAQTLGISAIAIILILILIVKFRIRALLTLVIASLLTALATGLPTGSIVND
                  ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



-337-

```

5  orf140-1      MDGWTQTLAQTLGISAIAIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND
   orf140ng-1.pep VLKKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFAPGVASLIF
   orf140-1      ILVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF
10  orf140ng-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLPPFALASVGAFSVMHVFLPPHPGPIAASEFYG
   orf140-1      GFPIFFDAGLIVMLPIVFATARRMKQDVLPPFALASIGAFSVMHVFLPPHPGPIAASEFYG
15  orf140ng-1.pep ANIGQVLILGLPTAFITWYFSGYMLGKVLGRAIHVPVPELLSGGTQDSDPPKEPAKAGTV
   orf140-1      ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTHVVPPELLSGGTQDNDLPKEPAKAGTV
20  orf140ng-1.pep RGESGSTLEKTVDGALAPACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC
   orf140-1      RGESGSALEKTVDGALAPVCSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC
25  orf140ng-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSEHNDSDG
   orf140-1      FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSEHNDSDG
   orf140ng-1.pep FWLVGRLLDMDVPTTLKTTWTVNQTLLIAFIGFALSALLFAIV
   orf140-1      FWLVGRLLDMDVPTTLKTTWTVNQTLLIALIGFALSALLFAIV

```

30 Furthermore, ORF140ng-1 is homologous to an *E. coli* protein:

```

35  gi|882633 (U29579) ORF_o454 [Escherichia coli] >gi|1789097 (AE000358) o454;
   This 454 aa ORF is 34% identical (9 gaps) to 444 residues of an approx. 456 aa
   protein GNTP BACLI SW: P46832 [Escherichia coli] Length = 454
   Score = 210 bits (529), Expect = 1e-53
   Identities = 130/384 (33%), Positives = 194/384 (49%), Gaps = 19/384 (4%)

40  Query: 88  ETSGGAQSLADALIRMFGEKRAPFAPGVASLIFGFPIFFDAGLIVMLPIVFATARRMKQD 147
   E SGA+SLA+ R G+KR A +A+ G P+FFD G I++ PI++ A+ K
   Sbjct: 80  EHSGGAESLANYFSRKLGDKRTIAALTIAAFFLGIPVFFDVGFIILAPIYGFYAKVAKIS 139

   Query: 148  VLPFALASVGAFSVMHVFLPPHPGPIAASEFYGANIGQVLILGLPTAFITWYFSGYMLGK 207
   L F L G +HV +PPHPGP+AA+ A+IG + I+G+ + I GY K
   Sbjct: 140  PLKFGLPVAGIMLTVHVAVPPHPGPVAAAGLLHADIGWLTIIIGIAIS-IPVGVVGYFAAK 198

45  Query: 208  VLGRAIHVPVPELL-----SGGTQDSDPPKEPAKAGTVVAVMLIPMLLIFLNTGV 257
   ++ + + E+L G T+ SD P A V ++++IP+ +I T
   Sbjct: 199  IINKRQYAMSVEVLEQMLAPASEEGATKLSKINPPGVA-LVTSLIVIPAIIMAGT-- 255

50  Query: 258  SALISEKLVSADETWVQTAKMIGSTPXXXXXXXXXXXXXGRKRGESGSTLEKTVDGALA 317
   +S L+ + T ++IGS +RG S + AL
   Sbjct: 256  ---VSATLMPPSHPLGLTLQIGSPMVALMIALVLAFWLLALRRGWSLQHTSDIMGSALP 312

   Query: 318  PACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGCFLVALALRIAQGSXXXX 377
   A VIL+TGAGG+FG VL SG+GKALA+ + + +P+L F+++LALR +QGS
55  Sbjct: 313  TAAVVILVTGAGGVFGKVLVESGVGKALANMLQMLDPLLPAAFIISLALRASQGS--AT 370

   Query: 378  XXXXXXXXXXXXXXXXGFTDWQLACIVLATAAGSVGCSEHNDSDGFWLVGRLLDMDVPTTLK 437
   G Q + LA G +G SH NDSEFW+V + L + V LK
60  Sbjct: 371  VAILTTGGLLSEAVMGLNPIQCVLVTLAACFGGLGASHINDSGFWIVTKYLGLSVADGLK 430

   Query: 438  TWTVNQTLLIAFIGFALSALLFAIV 461
   TWTV T++ F GF ++ ++A++
   Sbjct: 431  TWTVLTTILGFTGFLITWCVWAVI 454

```

Based on this analysis, including the identification of the presence of a putative leader sequence  
65 (double-underlined) and several putative transmembrane domains (single-underlined) in the

gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 71

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 593>:

```

5      1  ..GATTTCGGCA TATCGCCCGT GTATCTTTGG GTTGCCGCCG CGTTCAAACA
      51  TTTGCTGTGCG CCGTGGGCTG CCGACTCATA CGATGTCGCA CGCTTTGCAG
     101  GCGTATTTTTC TGCCTTTATC GGACTGACTT CCGTGGGCTT TGCCGGTTTC
     151  AACTTTTTCG GCAGACACCA CGGGCGCAC. GTCGTCCTGA TTCTCATCGG
     201  CTGTATCGGG CTGATTCCAG TTGCCCATTT CCTCAACCCC GCTGCCGCCG
     251  CCTTTGCCGC CGCCGGACTG GTGCTGCACG GTTATTCTTT GGCTCGCCCG
     301  CGCGTGATTG CCGCCTCTTT TCTGCTCGGT ACGGGCTGGA CGCTGATGTC
     351  GTTGGCAGCA GCTTATCCGG CAGCATTGTC CCTGATGCTG CCCTTGCCCG
     401  TACTGATGTT TTTCCGTCCG ..

```

This corresponds to the amino acid sequence <SEQ ID 594; ORF141>:

```

15      1  ..DFGISPVYLW VAAAFKHLIS PWAADSYDVA RFAGVFFAVI GLTSCGFAGF
      51  NFLGRHHGRX VVLILIGCIG LIPVAHFLNP AAAFAAAGL VLHGYSLARR
     101  RVIAASFLLG TGWTIMSLAA AYPAAFALML PLPVLMEFRP ..

```

Further work revealed the complete nucleotide sequence <SEQ ID 595>:

```

20      1  ATGCTGACCT ATACCCCGCC CGATGCCCGC CCGCCCGCCA AAACCCACGA
      51  AAAGCCGTGG CTGCTGCTGT TGATGGCGTT TGCCTGGTTG TGGCCCGGCG
     101  TGTTTTCCCA CGATTTGTGG AATCCTGACG AACCTGCCGT CTATACCGCC
     151  GTCGAAGCAC TGGCAGGCAG CCCACCCCC TTGGTTGCCG ATCTGTTCCG
     201  TCAAACCGAT TTCGGCATA CGCCCGTGTA TCTTTGGGTT GCCGCCGCGT
     251  TCAAACATTT GCTGTCGCCG TGGGCTGCCG ACTCATACGA TGCCGCACGC
     301  TTTGCAAGCG TATTTTTCG CGTTATCGGA CTGACTTCCT GCGGCTTTGC
     351  CGGTTTCAAC TTTTGGGCA GACACCACGG GCGCAGCGTC GTCCTGATTC
     401  TCATCGGCTG TATCGGCTG ATTCCAGTTG CCCATTTCTT CAACCCCGCT
     451  GCCGCCGCTT TTGCCCGCG CCGACTGGTG CTGCACGGTT ATTCTTTGGC
     501  TCGCCGGCGC GTGATTGCCG CCTCTTTTCT GCTCGGTACG GGCTGGACGC
     551  TGATGTGCTT GGCAGACGCT TATCCGGCAG CATTTGCCCT GATGCTGCCC
     601  TTGCCCGTAC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
     651  GACGGCAGTC GCCTCACTTG CCTTGCCCT GCCGTTATG ACCGTTTACC
     701  CGCTGCTCTT GGCAAAAACG CAGCCGCGC TGTTCCGCGA ATGGCTCGAC
     751  TATCACGTTT TCGGTACGTT CGGCGGCGTG CGGCACGTT AGACGGCATT
     801  CAGTTTGTCT TACTATCTGA AAAACCTGCT TTGGTTTGCA TTGCCCGCGC
     851  TGCCGCTGGC GGTTTGGACG GTTTGCCGCA CGCGCCTGTT TTCGACCGAC
     901  TGGGGGATTT TGGGCGTCGT CTGGATGCTT GCCGTTTGG TGCTGCTTGC
     951  CGTCAATCCG CAGCGTTTTC AGGATAACCT CGTCTGGCTG CTTCGCCCGC
    1001  TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGGCG CGGCGCGGCG
    1051  GCGTTTGTCA ACTGTTTCGG CATATGGCG TTCGGAAGT GTGCCGTGTT
    1101  CCTGTGGACG GGCTTTTTCG CCATGAATTA CGGCTGGCCC GCCAAGCTTG
    1151  CCGAACGCGC CGCCTATTTT AGCCCGTATT ATGTTCTCTA TATCGATCCC
    1201  ATTCCGATGG CGGTTGCCGT ACTGTTTACA CCCTTGTGGC TGTGGGCGAT
    1251  TACCCGGAAG AACATACGCG GCAGGCAGGC GGTACCAAC TGGGCGGCG
    1301  GCGTTACCCT GACCTGGGCT TTGCTGATGA CGCTGTTTCT GCCGTGGCTG
    1351  GACGCGGCGA AAAGCCACGC GCCGGTCGTC CGGAGTATGG AGGCATCGCT
    1401  TTCCCGGAA TTGAAACGGG AGCTTTCAGA CGGCATCGAG TGTATCGGCA
    1451  TAGGCGGCGG CGACCTGCAC ACGCGGATTG TTGGACGCA GTACGGCACA
    1501  TTGCCGCACC GCGTCGGCGA TGTACAATGC CGCTACCGCA TCGTCTCTCT
    1551  GCCCCAAAT GCGGATGCGC CGCAAGGCTG GCAGACGGTT TGGCAGGGTG
    1601  CGCGTCCGCG CAACAAAGAC AGTAAGTTCG CACTGATACG GAAAATCGGG
    1651  GAAAATATAT AA

```

This corresponds to the amino acid sequence <SEQ ID 596; ORF141-1>:

```

55      1  MLTYTPPDAR PPAKTHEKPW LLLLMAFAWL WPGVFSHDLW NPDEPAVYTA
      51  VEALAGSPTP LVAHLFGQTD FGIPPVYLWV AAFAKHLSP WAADSYDAAR
     101  FAGVFFAVIG LTSCGFAGFN FLGRHHGRSV VLILIGCIGL LIPVAHFLNPA
     151  AAAFAAAGLV LHGYSLARRR VIAASFLLGT GTWLMSLAAA YPAAAFALMLP
     201  LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLLAKT QPALFAQWLD

```

Computer analysis of this amino acid sequence gave the following results:

10 ORF141 shows 95.0% identity over a 140aa overlap with an ORF (ORF141a) from strain A of *N. meningitidis*:

The complete length ORF141a nucleotide sequence <SEQ ID 597> is:

	1	ATGCTGACCT	ATACCCCGCC	CGATGCCCGC	CCGCCCGCCA	AAACCCACGA
35	51	AAAGCCGTGG	CTGTTGCTGT	TGATGGCGTT	TGCCTGGTTG	TGGCCCGGCG
	101	TGTTTTCCCA	CGATTTGTGG	AATCCTGACG	AACCTGCCGT	CTATACCGCC
	151	GTCGAAGCAC	TGGCAGGCAG	CCCCACCCCT	TTGGTTGCCC	ATCTGTTTCGG
	201	TCAAATCGAT	TTCCGCATAC	CGCCCGTGTA	TCTTTGGGTT	CGCCGCGCGT
	251	TCAAACATTT	GCTGTCGCCG	TGGGCTGCCG	ACCCGTATGA	TGCCGCACGC
40	301	TTTGCCGGCG	TGTTTTTCGC	CGTTGTCGGA	CTGACTTCCT	GCGGCTTTGC
	351	CGGTTTCAAC	TTTTTGGGCA	GACACCACGG	GCGCAGCGTC	GTCCTGATTG
	401	TCATCGGCTG	TATCGGGCTG	ATTCCGACCG	TACACTTTCT	CAACCCCGCT
	451	CGCGCCGCCT	TTGCCGCGCG	CGGACTGGTG	CTGCACGGTT	ATTCTTTGGC
	501	TCGCCGGCGC	GTGATTGCCG	CCTCTTTTCT	GCTCGGTACG	GGTTGGACGC
45	551	TGATGTCGTT	GGCAGCAGCT	TATCCGGCGG	CATTTGCCCT	GATGCTGCCC
	601	CTGCCCGTGC	TGATGTTTTT	CCGTCCGTGG	CAAAGCAGCG	GTTTGATGTT
	651	GACGGCAGTC	GCCTCGCTTG	CCTTGCCCTT	GCGCCTTATG	ACCGTTTACC
	701	CGTGCTCTTT	GGCAAAAACG	CAGCCCGCGC	TGTTCCGCGA	ATGGCTCGAC
	751	GATCACGTTT	TCGGTACGTT	CGCGGCGGTG	CGGCACATTG	AGACGCGATT
50	801	CAGTTTGTTT	TACTATCTGA	AAAACCTGCT	TTGGTTTGCA	TTGCCCTGCGC
	851	TGCCCGTGCG	GGTTTGGACG	GTTTGGCCGA	CGCGCTCTGT	TTGACCCGAC
	901	TGGGGGATTT	TGGGCGTCGT	TGGGATGCTT	GCGGTTTTGG	TGCTGCTTGC
	951	CGTCAATCCG	CAGCGTTTTT	AGGATAACCT	CGTCTGGCTG	CTTCCGCCGC
	1001	TTGCCCTGTT	CGGCGCGGCG	CAACTGGACA	GCCTGAGACG	CGGCGCGGCG
55	1051	GCGTTTGTCA	AGTGTTTCGG	CATTATGGCG	TTCCGACTGT	TTGCCGTGTT
	1101	CGTGTGGACG	CCGTTTTTCG	CCATGAATTA	CGGCTAGGCT	GCCAAGCTTG
	1151	CCGAACGCGC	CGCCTATTTT	AGCCCGTATT	ATGTTCTCTG	TATCGATCCC
	1201	ATTCCGATGG	CGGTTGCCGT	ACTGTTTACA	CCCTTGTGGC	TGTGGGCGAT
	1251	TACCCGCAAA	AACATACGCG	GCAGGCAGCG	GGTTACCAAC	TGGGCGGCAG
	1301	GCGTTTACCT	GACCTGGGCT	TTCGCTATGA	CGCTTTCCTG	GCCGTGGCTG
60	1351	GACGCGGCGA	AAAGCCACGC	GCCGCTCGTC	CGGAGTATGG	AGGCATCGCT
	1401	TTCCCCGGAA	TTAAAACGGG	AGCTTTCAGA	CGGCATCGAG	TGTATCGACA
	1451	TAGCGCGGCG	CGACCTACAC	ACGCGGATTG	TTTGGACGCA	GTACGCGACA
	1501	TGCCCCGACC	GCGTCGGCGA	TGTACAATGC	CGCTACCGCA	TCGTCGCTTT
	1551	GCCCCAAAAC	CGGGATGCGC	CGCAAGGCTG	GCAGACGCTC	TGGCAGGGTG

-340-

1601 CGCGCCCGCG CAACAAAGAC AGTAAGTTCG CACTGATACG GAAAACCGGG  
 1651 GAAATATAT TAAAAACAAC AGATTGA

This encodes a protein having amino acid sequence <SEQ ID 598>:

5 1 MLTYTPPDAR PPAKTHEKWP LLLMAFAWL WPGVFSDHLW NPDEPAVYTA  
 51 VEALAGSPTP LVAHLFGQID FGIPPVYLWV AAFAKHLLSP WAADPYDAAR  
 101 FAGVFFAVVG LTSCGFAGFN FLGRHHGRSV VLILIGCIGL IPTVHFLNPA  
 151 AAFAAAAGLV LHGYSLARRR VIAASFLLGT GWTLMSLAAA YPAAFALMLP  
 201 LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLAKT QPALFAQWLD  
 251 DHVFGTFGGV RHIQTAFSLF YYLKNLLWFA LPALPLAVWT VCRTRLFSTD  
 10 301 WGILGVVWML AVLVLAVNP QRFQDNLVWL LPPLALFGAA QLDSLRRGAA  
 351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFY SPYYVPDIDP  
 401 IPMAVAVLFT PLWLWAITRK NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL  
 451 DAAKSHAPVV RSMEASLSPE LKRELSDGIE CIDIGGGDLH TRIVWTQYGT  
 501 LPHRVGDVQC RYRIVRLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKTG  
 15 551 ENILKTTD\*

ORF141a and ORF141-1 show 98.2% identity in 553 aa overlap:

orf141a.pep MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSDHLWNPDEPAVYTAVEALAGSPTP  
 |  
 orf141-1 MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSDHLWNPDEPAVYTAVEALAGSPTP  
 20 orf141a.pep LVAHLFGQIDFGIPPVYLWVAAAFKHLLSPWAADPYDAARFAGVFFAVVGLTSCGFAGFN  
 |  
 orf141-1 LVAHLFGQIDFGIPPVYLWVAAAFKHLLSPWAADSYDAARFAGVFFAVIGLTSCGFAGFN  
 25 orf141a.pep FLGRHHGRSVVLILIGCIGLIPTVHFLNPAFAAAGLV LHGYSLARRRVIAASFLLGT  
 |  
 orf141-1 FLGRHHGRSVVLILIGCIGLIPVAHFLNPAFAAAGLV LHGYSLARRRVIAASFLLGT  
 30 orf141a.pep GWTLMSLAAAYPAAAFALMLPLPVLMMFRPWQSRRLMLTAVASLAFALPLMTVYPLLAKT  
 |  
 orf141-1 GWTLMSLAAAYPAAAFALMLPLPVLMMFRPWQSRRLMLTAVASLAFALPLMTVYPLLAKT  
 orf141a.pep QPALFAQWLDLHVFGTFGGVRHIQTAFSLFY LKNLLWFALPALPLAVWTVCRTRLFSTD  
 |  
 35 orf141-1 QPALFAQWLDYHVFGTFGGVRHVQTAFSLFY LKNLLWFALPALPLAVWTVCRTRLFSTD  
 orf141a.pep WGILGVVWMLAVLVLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA  
 |  
 40 orf141-1 WGILGVVWMLAVLVLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA  
 orf141a.pep FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMVAVLFTPLWLWAITRK  
 |  
 orf141-1 FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMVAVLFTPLWLWAITRK  
 45 orf141a.pep NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSMELKRELSDGIE  
 |  
 orf141-1 NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSMELKRELSDGIE  
 orf141a.pep CIDIGGGDLHTRIVWTQYGTLP HRVGDVQC RYRIVRLPQNADAPQGWQTVWQGARPRNKD  
 |  
 50 orf141-1 CIDIGGGDLHTRIVWTQYGTLP HRVGDVQC RYRIVLLPQNADAPQGWQTVWQGARPRNKD  
 orf141a.pep SKFALIRKTGENI  
 |  
 55 orf141-1 SKFALIRKIGENI

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF141 shows 95% identity over a 140aa overlap with a predicted ORF (ORF141ng) from *N.gonorrhoeae*:

60 orf141.pep DFGISPVYLWVAAAFKHLLSPWAADSYDVA 30  
 |  
 orf141ng WNPAEPAVYTAVEALAGSPTPLVAHLFGQIDFGIPPVYLWVAAAFKHLLSPWAADPYDAA 126

-341-

	orf141.pep	RFAGVFFAVIGLTSCGFAGFNFLGRHHGRXVVLILIGCIGLIPVAHFLNPAAAAFAAAGL	90
	orf141ng	RFAGVFFAVIGLTSCGFAGFNFLGRHHGRSVVLIHIGCIGLIPVAHFFNPAAAAFAAAGL	186
5	orf141.pep	VLHGYSLARRRVIAASFLGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRP	140
	orf141ng	VLHGYSLARRRVIAASFLGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRMLMTA	246

An ORF141ng nucleotide sequence <SEQ ID 599> was predicted to encode a protein having amino acid sequence <SEQ ID 600>:

10	1	MPSEAVSARP	LCEYLLHLAI	RPFLTLMLT	YTPPDARPPA	KTHEKPWLLL
	51	LMAFAWLWPG	VFSHDLWNPA	EPAVYTAVEA	LAGSPTPLVA	HLFGQTDGFI
	101	PPVYLWVAAA	FKHLLSPWAA	HPYDAARFAG	VFFAVIGLTS	CGFAGFNFLG
	151	RHHGRSVVLI	HIGCIGLIPV	AHFFNPAAAA	FAAAGLVLHG	YSLARRRVIA
	201	ASFLGTGWT	LMSLAAAYPA	AFALMLPLPV	LMFFRPWQSR	RLMLTAVASL
15	251	AFALPLMTVY	PLLLAKTQPA	LFAQWLNHYV	FGTFGGVRHI	QRAFSLFHYL
	301	KNLLWFAPPG	LPLAVWTVCR	TRLFSTDWGI	LGIVWMLAVL	VLLAFNPQRF
	351	QDNLVWLLPP	LALFGAAQLD	SLRRGAAAFV	NWFGIMAFGL	FAVFLWTGFF
	401	AMNYGWPAKL	AERAAYFSPY	YVPDIDPIMP	AVAVLFTPLW	LWAITRKNIR
	451	GRQAVTNWAA	GVTLLTALLM	TLFLPWLDAA	KSHAPVVRSM	EASFSPELKR
20	501	ELSDGIECIG	IGGGDLHTRI	VWTQYGTLPH	RVGDVRCRYR	IVRLPQNADA
	551	PQGQTVWQG	ARPRNKDSKF	ALIRKIGENI	LKTTD*	

Further work revealed the following gonococcal DNA sequence <SEQ ID 601>:

	1	ATGCTGACCT	ATACCCCGCC	CGATGCCCGC	CCGCCCGCCA	AAACCCACGA
	51	AAAACCGTGG	CTGCTGCTGT	TGATGGCGTT	TGCCTGGCTG	TGGCCCGGCG
25	101	TGTTTTCCCA	CGATTTGTGG	AATCCTGCCG	AACCTGCCGT	CTATACCGCC
	151	GTCGAAGCAC	TGGCAGGCAG	CCCCACCCCC	TTGGTTGCCC	ATCTGTTCCG
	201	TCAAACCGAT	TTCGGCATA	CGCCCGTGTA	TCTTTGGGTT	GCCGCCGCAT
	251	TCAAACATTT	GCTGTCGCCG	TGGGCAGCCG	ACCCGTATGA	TGCCGCACGC
	301	TTTGCAGGCG	TATTTTTTGC	CGTTATCGGA	CTGACTTCTT	GCGGCTTTGC
30	351	CGGTTTCAAC	TTTTTGGGCA	GACACCACGG	GCGCAGCGTT	GTTTAAATCC
	401	ATATCGGCTG	TATCGGGCTG	ATTCCGGTTG	CCCATTTCCT	CAATCCcgcc
	451	gccgcgcgct	tTGCCGCCGC	CGGACTGGTG	CTGCacggct	actcgctgGC
	501	ACGCCGGCGC	GTGATtgcg	cctctTtccT	GCTCGGTACG	GGTTGGACGT
	551	TGATGTGCTG	GGCGGCAGCT	TATCCGGCGG	CGTTTGCCTG	GATGCTGCCC
35	601	CTGCCCCTGC	TGATGTTTTT	CCGTCCGTGG	CAAAGCAGGC	GTTTGATGTT
	651	GACGGCAGTC	GCCTCGCTTG	CCTTTGCCCT	GCCGCTTATG	ACCGTTTACC
	701	CGCTGCTCtt	gGCAAAAACG	CAGCCCGCGC	TGTTTGCCTG	ATGGCTCAAC
	751	TATCACGTTT	TCGGTACGtt	cggcgGCGTG	CGGCACaTTC	AGAggcatT
	801	Cagtttgttt	cactatctgA	AAaatctgct	ttggttcgca	ccgcccgggC
40	851	TGCCGCTGGC	GGTTTGGACG	GTTTGCCGCA	CACGCTGTG	TTCGACCGAC
	901	TGGGGGATTT	TGGGCATTGT	CTGGATGCTT	GCCGTTTGG	TGCTGCTCGC
	951	CTTTAATCCG	CAGCGTTTTC	AAGACAACCT	CGTCTGGCTG	CTGCCGCCGC
	1001	TTGCCCTGTT	CGGCCGCGCG	CAACTGGACA	GCCTGAGGCG	CGGCCGCGCG
	1051	GCTTTTGTCA	ACTGGTTCGG	CATTATGGCG	TTCGGGCTGT	TGCGCTGTT
45	1101	CCTGTGGACG	GGCTTTTTTC	CCATGAATTA	CGGCTGGCCC	GCCAAGCTTG
	1151	CCGAACGCGC	CGCCTACTTC	AGCCCGTATT	ACGTTCCCGA	CATCGATCCC
	1201	ATTCCGATGG	CGGTTGCCGT	ACTGTTTACA	CCCTTGTTGG	TGTGGGCGAT
	1251	TACCCGGAAA	AACATACCGG	GCAGGCAGGC	GGTTACCAAC	TGGGCGGCGAG
	1301	GCGTTACCCCT	GACCTGGGCT	TTGCTGATGA	CGCTGTTTCT	GCCGTGGCTG
50	1351	GACGCGGCGA	AAAGCCACGC	GCCCGTCGTC	CGGAGTATGG	AGGCATCGTT
	1401	TTCCCCGGAA	TAAAACGGG	AGCTTTCAGA	CGGCATCGAG	TGTATCGGCA
	1451	TAGGCGGCGG	CGACCTGCAC	ACGCGGATTG	TTTGACGCA	GTACGGCACA
	1501	TTGCCGCAAC	CGCTCGGCGA	TGTCCGTTGC	CGCTACCGTA	TCGTCGCCT
	1551	GCCCCAAAAC	GCGGATGCGC	CGCAAGGCTG	GCAGACGGTC	TGGCAGGGTG
55	1601	CGCGCCCGCG	CAACAAAGAC	AGTAAGTTTG	CACTGATACG	GAAAATCGGG
	1651	GAAAATATAT	TAAAAACAAC	AGATTGA		

This corresponds to the amino acid sequence <SEQ ID 602; ORF141ng-1>:

	1	MLTYTPPDAR	PPAKTHEKPW	LLLLMAFAWL	WPGVFSHDLW	NPAEPAVYTA
	51	VEALAGSPTP	LVAHLFGQTD	FGIPPVYLWV	AAAFKHLLSP	WAADPYDAAR
60	101	FAGVFFAVIG	LTSCGFAGFN	FLGRHHGRSV	VLIHIGCIGL	IPVAHFLNPA
	151	AAFAAAGLV	LHGYSLARRR	VIAASFLGT	GWTLMSLAAA	YPAAFALMLP
	201	LPVLMFFRPW	QSRRLMLTAV	ASLAFALPLM	TVYPLLLAKT	QPALFAQWLN
	251	YHVFGTGGV	RHIQRAFSLF	HYLKNLLWFA	PPGLPLAVWT	VCRLFLSTD
	301	WGILGIVWML	AVLVLLAFNP	QRFQDNLVWL	LPPLALFGAA	QLDSLRRGAA

-342-

```

351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFY SPYYVPDIDP
401 IPMAVAVLFT PLWLWAITRK NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL
451 DAAKSHAPVV RSMEASFSE LKRELSDGIE CIGIGGGDLH TRIVWTQYGT
501 LPHRVGDVRC RYRIVRLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKIG
551 ENILKTTD*

```

ORF141ng-1 and ORF141-1 show 97.5% identity in 553 aa overlap:

```

10 orf141ng-1.pep MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPAEPAVYTAVEALAGSPTP
    orf141-1      MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDPAVYTAVEALAGSPTP
    orf141ng-1.pep LVAHLEFGQTDGFI PPVYLWVAAAFKHLSPWAADPYDAARFAGVFFAVIGLTS CGFAGFN
    orf141-1      LVAHLEFGQTDGFI PPVYLWVAAAFKHLSPWAADSYDAARFAGVFFAVIGLTS CGFAGFN
    orf141ng-1.pep FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAAAAFAAAGLV LHGYS LARRRVIAAS FLLGT
    orf141-1      FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAAAAFAAAGLV LHGYS LARRRVIAAS FLLGT
    orf141ng-1.pep GWTLMSLAAAYPAAFALMLPLPVL MFFRPWQSRRMLTAVASLAFALPLMTVYPLLAKT
    orf141-1      GWTLMSLAAAYPAAFALMLPLPVL MFFRPWQSRRMLTAVASLAFALPLMTVYPLLAKT
    orf141ng-1.pep QPALFAQWLNHYHVF GTFGVVRHIQRAFS LFHYLKNLLWFAPPGLPLAVWTVCRTRLFSTD
    orf141-1      QPALFAQWLDYHVF GTFGVVRHVQTAFSLFY LKNLLWFALPALPLAVWTVCRTRLFSTD
    orf141ng-1.pep WGILGIVWMLAVLVLLAFNPQRQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
    orf141-1      WGILGVVWMLAVLVLLAVNPQRQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
    orf141ng-1.pep FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPM AVAVLFTPLWLWAITRK
    orf141-1      FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPM AVAVLFTPLWLWAITRK
    orf141ng-1.pep NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSM EASFSE LKRELSDGIE
    orf141-1      NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSM EASLSE LKRELSDGIE
    orf141ng-1.pep CIGIGGGDLHTRIVWTQYGTLP HRVGDVRCRYRIVRLPQNADAPQGWQTVWQGARPRNKD
    orf141-1      CIGIGGGDLHTRIVWTQYGTLP HRVGDVQCRYRIVLLPQNADAPQGWQTVWQGARPRNKD
    orf141ng-1.pep SKFALIRKIGENILKTTDX
    orf141-1      SKFALIRKIGENIX

```

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## Example 72

50 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 603>:

```

1  ..CAATCCGCCA AATGGTTATC GGGCCAACT CTAGTCGGCA CAGCAATTGG
51 GATACGCGGG CAGATAAAGC TTGGCGGCAA CCTGCATTAC GATATATTTA
101 CCGGCCGCGC ATTGAAAAAG CCCGAATTTT TCCAATCAAG GAAATGGGCA
151 AGCGGTTTTC AGGTAGGCTA TACGTTTTAA

```

55 This corresponds to the amino acid sequence <SEQ ID 604; ORF142>:

```

1  ..QSAKWLSGQT LVGTAIGIRG QIKLGGNLHY DIFTGRALKK PEFFQSRKWA
51 SGFQVGYTF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 605>:

-343-

1 ATGGATAATT CGGGTAGTGA GCGGACAGGA AAATACCAAG GAAATATCAC  
 51 TTTCTCTGCC GACAATCCTT TGGGACTGAG TGATATGTTT TATGTAAATT  
 101 ATGGACGTTC GATTGGCGGT ACGCCCGATG AGGAAAAGTTT TGACGGCCAT  
 151 CGCAAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCCTTT  
 5 C 201 CGGTAAATGG ACATGGGCAT TCAATCACAA TGGCTACCGT TACCATCAGG  
 251 CAGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAT  
 301 ACTGATTTTCG GCTTCAACCG CCTGTTGTAT CGTGATGCCA AACGCAAAAC  
 351 CTATCTCGGT GTAAAACTGT GGATGAGGGA AACAAAAAGT TACATTGATG  
 401 ATGCCGAAC TACTGTACAA CGGCGTAAAA CTGCGGGTTG GTTGGCAGAA  
 10 451 CTTTCCACACA AAGAATATAT CGGTCGCAGT ACGGCAGATT TTAAGTTGAA  
 501 ATATAAACGC GGCACCGGCA TGAAAGATGC TCTGCGCGCG CCTGAAGAAG  
 551 CCTTTGGCGA AGGCACGTCA CGTATGAAAA TTTGGACGGC ATCGGCTGAT  
 601 GTAAATACTC CTTTTCAAAT CGGTAAACAG CTATTTGCCT ATGACACATC  
 651 CGTTTCATGCA CAATGGAACA AAACCCCGCT AACATCGCAA GACAACTGG  
 15 701 CTATCGGCGG ACACCAACC GTACGTGGCT TCGACGGTGA AATGAGTTTG  
 751 TCTGCCGAGC GGGGATGGTA TTGGCGCAAC GATTGAGCT GGCAATTTAA  
 801 ACCAGGCCAT CAGCTTTATC TTGGGGCTGA TGAGGACAT GTTTCAGGAC  
 851 AATCCGCCAA ATGGTTATCG GGCCAAACTC TAGTCGGCAC AGCAATTGGG  
 901 ATACGCGGGC AGATAAAGCT TGGCGGCAAC CTGCATTACG ATATATTTAC  
 20 951 CGGCCGCGCA TTGAAAAAGC CCGAATTTT CCAATCAAGG AAATGGGCAA  
 1001 GCGGTTTTC GGTAGCTAT ACGTTTAA

This corresponds to the amino acid sequence <SEQ ID 606; ORF142-1>:

1 MDNSGSEATG KYQGNITFSA DNPLGLSDMF YVNYGRSIGG TPDEESFDGH  
 25 51 RKEGGSNNYA VHYSAPFGKW TWAFNHNGYR YHQAVSGLSE VYDYNKSYN  
 101 TDFGFNRLLY RDAKRRTYLG VKLWMRETKS YIDDAELTVQ RRKTAGWLAE  
 151 LSHKEYIGRS TADFKLKYKR GTGMKDALRA PEEAFGEGTS RMKIWTASAD  
 201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL  
 251 SAERGWYWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWLS GQTLVGTAIG  
 301 IRGQIKLGGN LHYDIFTGRA LKKPEFFQSR KWASGFQVGY TF\*

30 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF142 shows 88.1% identity over a 59aa overlap with a predicted ORF (ORF142ng) from *N.gonorrhoeae*:

35 orf142.pep QSAKWLSGQTLVGTAIGIRGQIKLGGNLHY 30  
 orf142ng RGWYWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLAGTAIGIRGQIKLGGNLHY 313  
 orf142.pep DIFTGRALKKPEFFQSRKWASGFQVGYTF 59  
 40 orf142ng DIFTGRALKKPEYFQTKKWTGFQVGYSF 342

The complete length ORF142ng nucleotide sequence <SEQ ID 607> is:

1 ATGGATAATT CGGGTAGTGA GCGGACAGGA AAATACCAAG GAAATATCAC  
 51 TTTCTCTGCC GACAATCCTT TTGGACTGAG TGATATGTTT TATGTAAATT  
 45 101 ATGGACGTTC AATTGGCGGT ACGCCCGATG AGGAAAATTT TGACGGCCAT  
 151 CGCAAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCCTTT  
 201 CGGTAAATGG ACATGGGCAT TCAATCACAA TGGCTACCGT TACCATCAGG  
 251 CGGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAC  
 301 ACTGATTTTCG GCTTCAACCG CCTGTTGTAT CGTGATGCCA AACGCAAAAC  
 351 CTATCTCAGT GTAAAACTGT GGACGAGGGA AACAAAAAGT TACATTGATG  
 50 401 ATGCCGAAC TACTGTACAA CGGCGTAAAA CCACAGGTTG GTTGGCAGAA  
 451 CTTTCCACACA AAGGATATAT CGGTCGCAGT ACGGCAGATT TTAAGTTGAA  
 501 ATATAACAC GGCACCGGCA TGAAAGATGC TCTGCGCGCG CCTGAAGAAG  
 551 CCTTTGGCGA AGGCACGTCA CGTATGAAAA TTTGGACGGC ATCGGCTGAT  
 601 GTAAATACTC CTTTTCAAAT CGGTAAACAG CTATTTGCCT ATGACACATC  
 55 651 CGTTTCATGCA CAATGGAACA AAACCCCGCT AACATCGCAA GACAACTGG  
 701 CTATCGGCGG ACACCAACC GTACGTGGCT TCGACGGTGA AATGAGTTTG  
 751 CCTGCCGAGC GGGGATGGTA TTGGCGCAAC GATTGAGCT GGCAATTTAA  
 801 ACCAGGCCAT CAGCTTTATC TTGGGGCTGA TGAGGACAT GTTTCAGGAC  
 851 AATCCGCCAA ATGGTTATCG GGCCAAACTC TAGCCGGCAC AGCAATTGGG  
 60 901 ATACGCGGGC AGATAAAGCT TGGCGGCAAC CTGCATTACG ATATATTTAC  
 951 CGGCCGTGCA TTGAAAAAGC CCGAATATTT TCAGACGAAG AAATGGGTAA

-344-

1001 CGGGGTTTCA GGTGGGTTAT TCGTTTTGA

This encodes a protein having amino acid sequence &lt;SEQ ID 608&gt;:

5  
1 MDNSGSEATG KYQGNITFSA DNPFGLSDMF YVNYGRSIGG TPDEENFDGH  
51 RKEGGSNNYA VHYSAPFGKW TWA FNHNGYR YHQA VSGLSE VYDNGKSYN  
101 TDFGNRLLY RDAKRKTYLS VKLWTRETKS YIDDAELTVQ RRKTTGWLAE  
151 LSHKGYIGRS TADFKLKYKH GTGMKDALRA PEEAFGEGTS RMKIWTASAD  
201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL  
251 PAERGWYWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWLS GQTLAGTAIG  
301 IRGQIKLGGN LHYDIFTGRA LKKPEYFQTK KWTGTFQVGY SF\*

10 The underlined sequence (aromatic-Xaa-aromatic amino acid motif) is usually found at the C-terminal end of outer membrane proteins.

ORF142ng and ORF142-1 show 95.6% identity over 342aa overlap:

15  
orfl42-1.pep MDNSGSEATGKYQGNITFSADNPLGLSDMFYVNYGRSIGGTPDEESFDGHRKEGGSNNYA  
orfl42ng-1 MDNSGSEATGKYQGNITFSADNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYA  
20  
orfl42-1.pep VHYSAPFGKWTWAFNHNHNGYRYHQA VSGLSEVYDNGKSYNTDFGNRLLYRDAKRKTYLG  
orfl42ng-1 VHYSAPFGKWTWAFNHNHNGYRYHQA VSGLSEVYDNGKSYNTDFGNRLLYRDAKRKTYLS  
25  
orfl42-1.pep VKLWMRETKSYIDDAELTVQRRKTAGWLAELSHKEYIGRSTADFKLKYKRGTMKDALRA  
orfl42ng-1 VKLWTRETKSYIDDAELTVQRRKTGWLAELSHKGYIGRSTADFKLKYKHGTGMKDALRA  
30  
orfl42-1.pep PEEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT  
orfl42ng-1 PEEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT  
35  
orfl42-1.pep VRGFDGEMSLSAERGWYWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLVGTAG  
orfl42ng-1 VRGFDGEMSLPAERGWYWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLAGTAIG  
orfl42-1.pep IRGQIKLGGNLHYDIFTGRALKKPEFFQSRKWASGFQVGYTF  
orfl42ng-1 IRGQIKLGGNLHYDIFTGRALKKPEYFQTKKWTGTFQVGYSF

In addition, ORF142ng is homologous to the HecB protein of *E. chrysanthemi*:

40  
gi|1772622 (L39897) HecB [Erwinia chrysanthemi] Length = 558  
Score = 119 bits (295), Expect = 3e-26  
Identities = 88/346 (25%), Positives = 151/346 (43%), Gaps = 22/346 (6%)  
Query: 2 DMSGSEATGKYQGNITFSADNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYAV 61  
DMSG ++TG+ Q N + + DN FGL+D ++++ G S + + D + G  
Sbjct: 230 DMSGQKSTGEEQLNGSLALDNVFLADQWFISAGHS---SRFATSHDAESLQAG----- 280  
45  
Query: 62 HYSAPFGKWTWAFNHNHNGYRYHQA VSGLSEVYDNGKSYNTDFGNRLLYRDAKRKTYLSV 121  
+S P+G W +N++ RY + G S F +R+++RD KT ++  
Sbjct: 281 -FSMPYGYWNLGYNYSQSRYRNTFINRDFPWHSTGSDTHRFSLSRVVRDGTMTAIAIAG 339  
50  
Query: 122 KLWTRETKSYIDDAELTVQRRKTGWLAELSHKGYIGRSTADFKLKYKHGTGMKDALARAP 181  
R +Y++ + L RK + ++H + A F Y G +  
Sbjct: 340 TFSQRTGNYYLNGSLPSSSRKLSSVSLGVNHSQKLWGGGLATFNPTYNRGVRWLGETDT 399  
55  
Query: 182 EEAFFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHTV 241  
+++ E + WT SA P Y S++ Q++ L ++L +GG ++  
Sbjct: 400 DKSADPEPRAEFNKWTLASASYHPV---TDSITYLGSGLYGYQSARALYGSEQLTLGGESSI 456  
60  
Query: 242 RGFDGEMSLPAERGWYWRNDLSWQFKP----GHQLYLGA-DVGHVSGQSAKWLSGQTLAG 296  
RGF E RG YWRN+L+WQ G+ ++ A D GH+ + +L G  
Sbjct: 457 RGF-REQYTSNGRGAYWRNELNWQAWQLPVLGNVTFMAAVDGGHLYNHKQDNSTAASLWG 515  
Query: 297 TAIGIRGQIKLGGNLHYDIFTGRALKKPEYFQTKKWTGTFQVGYYSF 342  
A+G+ + L + G + P + Q V G++VG SF



Sbjct: 516 GAVGMTVASRW---LSQQVTVGWPIISYPALWLPDPTMVVGYRVGLSF 558

On the basis of this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 5 Example 73

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 609>:

```

1  ATGCGGACGA AATGGTCAGC AGTGAGAAGC TGCTTACTTG GgCGGACACC
51  GCCGACATCG ATACCGCTTT GAACCTGTTG TACCGTTTGC AAAAATCGA
101 ATTCTCTAT  GGCGATGAAA ACGGTCATTC AGACGGCATC AATTGwCGG
151 ACGAGCAATT  GCCGTTGCTG ATGGAACAAT TGTCGGCAG  CGGTAAGGCG
201 TTATTGGTCG ATCGGAACGG TCTGTATCTT GCCAACGCCA ATTTCCATCA
251 TGAGGCGGCG GAAGAGTTGG GGTGTGTGGC GGCAGAAGTC GCACAGATGG
301 AAAAGAAATA CCGGCTGCTG ATTAAGAACA AC..

```

This corresponds to the amino acid sequence <SEQ ID 610; ORF143>:

```

15  1  MRTKWSAVRS CTWADTADID TALNLLYRLQ KLEFLYGDEN GHSDGINLXD
    51  EQPLLMQQL  SGSGKALLVD RNgLYLANAN FHHEAAEELG LLAAEVAQME
    101  KKYRLLIKNN ..

```

Further work revealed the complete nucleotide sequence <SEQ ID 611>:

```

20  1  ATGGAATCAA CACTTTCCTT ACAAGCAAAT TTATATCCCC GCCTGACTCC
    51  TGCCGGTGCA TTTTATGCCG TATCCAGCGA TGCCCCCAGT GCCCGTAAAA
    101  CTTTGTGCGA CAGCCTGTTG AAAGCAGATG CCGACGAAAT GGTCAGCAGT
    151  GAGAAGCTGC TTACTTGGGC GGACACCGCC GACATCGATA CCGCTTTGAA
    201  CCGTTGTGAC CGTTTGCAAA AACTCGAATT CCTCTATGGC GATGAAAACG
    251  GTCATTGAGA CCGCATCAAT TTGTCGGACG AGCAATTGCC GTTGCTGATG
    301  GAACAATTGT CCGGCAGCGG TAAGGCGTTA TTGGTCGATC GGAACGGTCT
    351  GTATCTTGCC AACGCCAATT TCCATCATGA GGCGGCGGAA GAGTTGGGGT
    401  TGTTGGCGGC AGAAGTCGCA CAGATGGAAA AGAAATACCG GCTGCTGATT
    451  AAGAACAACC TGTATATCAA CAATAACGCT TGGGGCGTTT GCGATCCTTC
    501  CGGTCAGAGC GAATTGACAT TTTTCCCATT GTATATCGGT TCAACCAAAT
    551  TTATTTTGGT TATCGGCGGC ATTCCCGATT TGGGCAAAGA GGCATTGTGT
    601  ACTTTGTGTA GGATTTTATA CCGCCGTTAC AGCAACCGCG TGTA

```

This corresponds to the amino acid sequence <SEQ ID 612; ORF143-1>:

```

35  1  MESTLSLQAN LYPRLTGAGA FYAVSSDAPS AGKTLHSLK KADADEMVSS
    51  EKLLTWADTA DIDTALNLLY RLQKLEFLYG DENGHS DGIN LSDEQLPLLM
    101  EQLSGSGKAL LVDRNGLYLA NANFHHEAAE ELGLLAAEVA QMEKKYRLLI
    151  KNNLYINNNA WGVCDPSGQS ELTFFPLYIG STKFILVIGG IPDLGKEAFV
    201  TLVRILYRRY SNRV*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

40 ORF143 shows 92.4% identity over a 105aa overlap with an ORF (ORF143a) from strain A of *N.meningitidis*:

```

                                     10      20      30
or:143.pep                        MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFL
                                     | : : | | | | | | | | | | | | | | | |
45  orf143a      GAFYAVSSDXPSAGKTLHSLKADADEMVSSKLLTWAXTADIDTALNLLYRLQKLEFL
                   20      30      40      50      60      70
                                     40      50      60      70      80      90
or:143.pep      YGDENGHS DGINLXDEQLPLLMQQLSGSGKALLVDRNGLYLANANFHHEAAEELGLLAAE
                   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

-346-

```

orf143a      YGDENGHSDGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLANANFHHEAAEELGLLAAE
              80          90          100          110          120          130
5            100          110
orf143.pep   VAQMEKKYRLLIKNN
              |||||
orf143a      VAQMEKKYRLXIKNNLYINNNNAWGVCDPGQSELTFFFLYIGSTKFIIVIGGIPDLGKEA
              140          150          160          170          180          190

```

The complete length ORF143a nucleotide sequence <SEQ ID 613> is:

```

10      1  ATGGAATCAA CANTTTCCT ACAAGCAAAT TTATATCNCC GCCTGACTCC
        51  TGCCGGTGCA TTTTATGCCG TATCCAGCGA TGNCCCCAGT GCCGGTAAAA
       101  CTTTGTTCGA CAGCCTGTTG AAAGCGGATG CGGACGAAAT GGTNAGCAGT
       151  GAGAAGCTGC TTACCTGGGC GGANACCGCC GACATCGATA CCGCTTTGAA
       201  CCTGTTGTAC CGTTTGCAAA AACTCGAATT CCTCTATGGC GATGAAAACG
15      251  CTCATTGAGA CGGCATCAAT TTGTCGGACG AGCAATTGCC GTTGCTGATG
       301  GAACAATTGT CCGGCAGCGG TAAGGCGTTA TTGGTCGATC GGAACGGTCT
       351  GTATCTTGCC AACGCCAATT TCCATCATGA GCGGCGGAA GAGTTGGGGT
       401  TGTTGGCGGC AGAAGTCGCA CAGATGGAAA AGAAATACCG GCTGCNNATT
       451  AAGAACAACC TGTATATCAA CAATAACGCT TGGGCGGTTT GCGATCCTTC
20      501  CGGTCAGAGC GAATTGACAT TTTTCCCAT GTATATCGGT TCAACCAAAT
       551  TTATTTTGGT TATCGGCGGC ATTCCCGATT TGGGCAAAGA GGCATTTGTT
       601  ACTTTGGTAA GGATNTTATA CCNCCNGTTA CAGCAACCGC GTGTAAACT
       651  TGGGAGAGAG GANGGGTTAT GCAGCAATTA TTGA

```

This encodes a protein having amino acid sequence <SEQ ID 614>:

```

25      1  MESTXSLQAN LYXRLTPAGA FYAVSSDXPS AGKTLHLSLL KADADEMVSS
        51  EKLLTWAXTA DIDTALNLLY RLQKLEFLYG DENGHSDGIN LSDEQLPLLM
       101  EQLSGSGKAL LVDRNGLYLA NANFHHEAAE ELGLLAAEVA QMEKKYRLXI
       151  KNNLYINNNA WGVCDPSGQS ELTFFFLYIG STKFILVIGG IPDLGKEAFV
       201  TLVRXLYXXL QQPRVKLGRE XGLCSNY*

```

30 ORF143a and ORF143-1 show 97.1% identity in 207 aa overlap:

```

        orf143a.pep   MESTXSLQANLYXRLTPAGAFYAVSSDXPSAGKTLHLSLLKADADEMVSSEKLLTWAXTA
        orf143-1      MESTLSLQANLYPRLTPAGAFYAVSSDAPSAGKTLHLSLLKADADEMVSSEKLLTWADTA
35      orf143a.pep   DIDTALNLLYRLQKLEFLYG DENGHSDGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA
        orf143-1      DIDTALNLLYRLQKLEFLYG DENGHSDGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA
        orf143a.pep   NANFHHEAAEELGLLAAEVAQMEKKYRLXIKNNLYINNNNAWGVCDPGQSELTFFFLYIG
        orf143-1      NANFHHEAAEELGLLAAEVAQMEKKYRLLIKNNLYINNNNAWGVCDPGQSELTFFFLYIG
        orf143a.pep   STKFILVIGGIPDLGKEAFVTLVRXLY
        orf143-1      STKFILVIGGIPDLGKEAFVTLVRILY
45      orf143-1      STKFILVIGGIPDLGKEAFVTLVRILY

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF143 shows 95.5% identity over a 110aa overlap with a predicted ORF (ORF143ng) from *N.gonorrhoeae*:

```

50      orf143.pep   MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFLYG DENGHSDGINLXDEQLPLLMEQL 60
        orf143ng     MRTKWSAVRSCSRADTADIDTALNLLYRLQKLEFLYG DENGHSDGINLSDEQLPLLMEQL 60
55      orf143.pep   SGSGKALLVDRNGLYLANANFHHEAAEELGLLAAEVAQMEKKYRLLIKNN 110
        orf143ng     SGSGKALLVDRNGLYLANANFHHEAAEELGLLAAEVAQMEKKYRLIRNNLYINNNNAWGV 120

```

An ORF143ng nucleotide sequence <SEQ ID 615> was predicted to encode a protein having amino acid sequence <SEQ ID 616>:

Further work revealed the following gonococcal DNA sequence <SEQ ID 617>:

15

20

This corresponds to the amino acid sequence <SEQ ID 618; ORF143ng-1>:

25

ORF143ng-1 and ORF143-1 show 95.8% identity in 214 aa overlap:

30

35

40

Based on the presence of the putative transmembrane domains<sup>5</sup> in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 74

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 619>:

50

55

```

1   ATGACCTTTT   TACAACGTTT   GCAAGGTTTG   GCAGACAATA   AAATCTGTGC
51  GTTTCATGCG   TTCGTCGTCC   GCCGCTTTGA   TGAAGAACGC   GTACCGCAGr
101 CGCGCGGCAAG   CATGACGTTT   ACGACGCTGC   TGGCACTCGT   CCCCGTGTCT
151 ACCGTGATAGG   TTGCGGTTCG   TTCGATTCTT   CCCGTGTTCG   ACCGTGTGTC
201 GGATTTCGTT   GTCTCCTTCG   TCAACCAAAC   CATTGTGCGC   CA .GGCGCGG
251 ACATGGTGTT   CGACTATATC   AATGCGTTCC   GCGAGCAGGC   GAACCGGCTG
301 ACGGCAATCG   GCAGCGTGAT   GCTGGTCGTT   ACCTCGTGTA   TGCTGATTCT
351 GACGATAGAC   AATACGTTCA   ACCGCATCTG   GgCGGGTCAA   wTyCCAGCGT
401 CCGTGGATG .

```

This corresponds to the amino acid sequence <SEQ ID 620; ORF144>:

```

1  MTFLLRLQGL ADNKAFAW FVRRFDEER VPQXAASMTF TLLALVPVL
51  TVMVAVASIF PVFDRWDSF VSFVNQITVP XGADMVFDYI NAFREQANRL
101 TAIGSVMLV TSLMLIRTID NTFNRIWRVX XQRPWM...

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 621>:

```

1  ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
51  GTTTGCATGG TTCGTCGTCC GCCGCTTGA TGAAGAACGC GTACCGCAGG
101 CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCCCTGTCTG
151 ACCGTGATGG TGGCGGTCGC TTCGATTTTC CCGGTGTTTC ACCGCTGGTC
201 GGATTCGTTT GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGCGCGCG
251 ACATGGTGTG CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTCTG
351 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCAGCGCTC
401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCCG
15 451 CTGTCTTTGG GCGTGGGCAT TTCTTTATG GTCGGCTCGG TACAGGATGC
501 CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGGCGG
551 CGACGCTGAC CTTTCATGAC CTTTGTCTGT GGGGCTGTA CCGCTTCGTG
601 CCAAACCGCT TCGTTCGCCG GCGGCAGGCG TTTGTCGGGG CTTTGGCAAC
651 AGCGTTTTGT CTGGAACCGC CGCGCTCCCT CTTCACTTGG TATATGGGCA
20 701 ATTTGACGCG CTACCGCTCG ATTTACGGCG CGTTGCGCG CGTGCCGTTT
751 TTTCTGTGTG GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT
801 GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGGGCT
851 TCGACTCGCG CGGACGGTTT GACGACGTGT TGAAATCCT GCTGCTTCTG
901 GATGCGGCGC AAAAAGAAGG CAAAGCCTTG CCTGTTCAGG AGTTCAGACG
25 951 GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
1001 CGCGGCACGG CTACATCTAT TCCGGCAGAC AGGGTTGGGT GTTGAAAACG
1051 GGGGCGGATT CGATTGAGTT GAACGAATC TTCAAGCTCT TCGTTTACCG
1101 TCCGTTGCCG GTGGAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
1151 TGACACCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTGACGCT
30 1201 CAGGCGAAAA AACGGCAGTA G

```

This corresponds to the amino acid sequence <SEQ ID 622; ORF144-1>:

```

1  MTFLLRLQGL ADNKAFAW FVRRFDEER VPQXAASMTF TLLALVPVL
51  TVMVAVASIF PVFDRWDSF VSFVNQITVP QGADMVFDYI NAFREQANRL
101 TAIGSVMLV TSLMLIRTID NTFNRIWRVN SQRPMWQFL VYWALLTFGP
35 151 LSLGVGISFM VGSVQDAALA SGAPQWSGAL RTAATLTFMT LLLWGLYRFV
201 PNRFPARQA FVGALATAFC LETARSLFTW YMGNEFDGYRS IYGAFAAVPF
251 FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RRGFDSRGRF DDVLKILLLL
301 DAAQKEGKAL PVQEFRRHIN MGYDELGELL EKLARHGYIY SGRQGWVLKT
351 GADSIELNEL FKLFVYRPLP VERDHVNQAV DAVMTPCLQT LNMTLAEFDA
40 401 QAKKRQ*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF144 shows 96.3% identity over a 136aa overlap with an ORF (ORF144a) from strain A of *N. meningitidis*:

```

45      10      20      30      40      50      60
orfl44.pep MTFLLRLQGLADNKAFAWFVRRFDEERVVPQXAASMTFTLLALVPVLTVMVAVASIF
          |||
orfl44a    MTFLLRLQGLADNKAFAWFVRRFDEERVVPQXAASMTFTLLALVPVLTVMVAVASIF
          |||
50      70      80      90     100     110     120
orfl44.pep PVFDRWDSFVSVFNQITVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
          |||
orfl44a    PVFDRWDSFVSVFNQITVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
          |||
55      70      80      90     100     110     120
          130
orfl44.pep NTFNRIWRVXXQRPWM
          |||
60 orfl44a    NTFNRIWRVNSQRPMWQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL

```

This corresponds to the amino acid sequence <SEQ ID 620; ORF144>:

```

1  MTFLLQRLQGL ADNKICAFW FVRRFDEER VPQXAASMTF TLLALVPVL
51  TVMVAVASIF PVFDRWSDSF VSFVNQTIVP XGADMVFDYI NAFREQANRL
101 TAIGSVMLV TSLMLIRTID NTFNRIWRVX XQRPWM...

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 621>:

```

1  ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
51  GTTTCATGG TTCGTCGTCC GCCGCTTGA TGAAGAACGC GTACCGCAGG
101 CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCCCTGTCTG
151 ACCGTGATGG TGGCGGTCGC TTCGATTTTC CCGGTGTTCG ACCGCTGGTC
201 GGATTCGTTT GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGCGCGCG
251 ACATGGTGTG CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTCTG
351 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCAGCGCTC
401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCCG
15  451 CTGTCTTTGG GCGTGGGCAT TTCCTTTATG GTCGGCTCGG TACAGGATGC
501 CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGGCGG
551 CGACGCTGAC CTTTCATGAC CTTTGTCTGT GGGGCTGTA CCGCTTCGTG
601 CCAAACCGCT TCGTTCCTCG GCGGCGAGCG TTTGTCGGGG CTTGGGCAAC
651 AGCGTTTTGT CTGGAACCGC CGCGCTCCCT CTTCACTTGG TATATGGGCA
20  701 ATTTGACGGC CTACCGTTCG ATTTACGGCG CGTTGCGCG CGTGCCGTTT
751 TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT
801 GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGGGCT
851 TCGACTCGCG CGGACGTTT GACGACGTGT TGAAAATCCT GCTGCTTCTG
901 GATGCGGCGC AAAAAGAAGG CAAAGCCTTG CCTGTTCAGG AGTTCAGACG
25  951 GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
1001 CGCGGCACGG CTACATCTAT TCCGGCAGAC AGGGTTGGGT GTTGAAAACG
1051 GGGGCGGATT CGATTGAGTT GAACGAATC TTCAAGCTCT TCGTTTACCG
1101 TCCGTTGCCT GTGGAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
1151 TGACACCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTGACGCT
30  1201 CAGGCGAAAA AACGGCAGTA G

```

This corresponds to the amino acid sequence <SEQ ID 622; ORF144-1>:

```

1  MTFLLQRLQGL ADNKICAFW FVRRFDEER VPQAAASMTF TLLALVPVL
51  TVMVAVASIF PVFDRWSDSF VSFVNQTIVP QGADMVFDYI NAFREQANRL
101 TAIGSVMLV TSLMLIRTID NTFNRIWRVN SQRPWMMQFL VYWALLTFGP
35  151 LSLGVGISFM VGSVQDAALA SGAPQWSGAL RTAATLTFT LLLWGLYRFV
201 PNRFPARQA FVGALATAFC LETARSLFTW YMGNFQGYRS IYGAFAAVPF
251 FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RRGFDSRGRF DDVLKILLLL
301 DAAQKEGKAL PVQEFRRHIN MGYDELGELL EKLARHGYIY SGRQGWVLKT
351 GADSIELNEL FKLFVYRPLP VERDHVNQAV DAVMTPCLQT LNMTLAEFDA
40  401 QAKKRQ*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF144 shows 96.3% identity over a 136aa overlap with an ORF (ORF144a) from strain A of *N. meningitidis*:

```

45  orf144.pep 10 20 30 40 50 60
      MTFLLQRLQGLADNKICAFW FVRRFDEERVPQXAASMTFTLLALVPVLTVMVAVASIF
      |||
orf144a MTFLLQRLQGLADNKICAFW FVRRFDEERVPQXAASMTFTLLALVPVLTVMVAVASIF
      |||
50  orf144.pep 70 80 90 100 110 120
      PVFDRWSDSFVSFVNQTIVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
      |||
orf144a PVFDRWSDSFVSFVNQTIVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
      |||
55  orf144.pep 70 80 90 100 110 120
      NTFNRIWRVXXQRPWM
      |||
60  orf144a NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL

```

130 140 150 160 170 180

The complete length ORF144a nucleotide sequence <SEQ ID 623> is:

```

1   ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
5   51  GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG
    101  CGGCGGCAAG CATGACGTTT ACGACACTGC TGGCACTCGT CCCCCTGTCTG
    151  ACCGTGATGG TGGCGGTCGC TTCGATTTTC CCCGTGTTCG ACCGNTGGTC
    201  GGATTCGTTT GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGGCGCGG
    251  ACATGGTNTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
    301  ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCNGA TGCTGATTCG
10  351  GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCCAGCGTC
    401  CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCCG
    451  CTGTCTTTGG GCGTGGGCAT TTCCTTTATN GTCGGCTCGG TACAGGATGC
    501  CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGGCGG
    551  CGACGCTGAN CTTTCATGAC CTTTTGCTGT GGGGGCTGTA CCGCTNCGTG
15  601  CCAAACCGCT TCGTTCGCCG GCGGCANGCG TTTGTGCGGG CTTTGGCAAC
    651  AGCGTTCTGT CTGGAACCG CGCGTTCCTT CTTTACTTGG TATATGGGCA
    701  ATTTTCGACG CTACCGCTCG ATTTACGNG CGTTTGCCGC CGTGCCGTTT
    751  TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT
    801  GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGNGCT
20  851  TCGACTCGCG CGGACGGTTT GACGACGTGT TGAAAATCCT GCTGCTTCTG
    901  GATGCGGCGC AAAAAGAAG CNAAGCCTTG CCTGTTCAGG AGTTCAGACG
    951  GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
    1001 CGCGGCACGG CTACATCTAT TCCGGCAGAC AGGGTTGGGT GTTGAAAACG
    1051 GGGGCGGATT CGATTGAGTT GAACGAATC TTCAAGCTCT TCGTTTACCG
25  1101 TCCGTTGCCT GTGGAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
    1151 TGATGCCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTGACGCT
    1201 CAGGCGAAAA AACAGCAGCA ATCTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 624>:

```

1   MTFLLRLQGL ADNKICAFW FVRRFDEER VPQAAASMTF TLLALVPVL
30  51  TVMVAVASIF PVFDRWSDSF VSFVNQTIVP QGADMVFDYI NAFREQANRL
    101  TAIGSVMLVV TSXMLIRTID NTFNRIWRVN SQRPMWQFL VYWALLTFGP
    151  LSLGVGISFX VGSVQDAALA SGAPQWSGAL RTAATLXFMT LLLWGLYRXV
    201  PNRFPARXA FVGALATAFC LETARSLFTW YMGNFDGYRS IYGAFAAVPF
35  251  FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RRXFDSRGRF DDVLKILLLL
    301  DAAQKEGXAL PVQEFRRHIN MGYDELGELL EKLRHGYIY SGRQGWVLKT
    351  GADSIENEL FKL FVYRPLP VERDHNQAV DAVMMPCLOT LNMTLAEFDA
    401  QAKKQQQS*

```

ORF144a and ORF144-1 show 97.8% identity in 406 aa overlap:

```

40  orf144a.pep  MTFLLRLQGLADNKICAFW FVRRFDEER VPQAAASMTFTLLALVPVLTVMVAVASIF
    orf144-1     MTFLLRLQGLADNKICAFW FVRRFDEER VPQAAASMTFTLLALVPVLTVMVAVASIF

    orf144a.pep  PVFDRWSDSFVSFVNQTIVPQADMVFDYINAFREQANRLTAIGSVMLVVT SXMLIRTID
45  orf144-1     PVFDRWSDSFVSFVNQTIVPQADMVFDYINAFREQANRLTAIGSVMLVVT SLMLIRTID

    orf144a.pep  NTFNRIWRVNSQRPMWQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL
50  orf144-1     NTFNRIWRVNSQRPMWQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL

    orf144a.pep  RTAATLXFMTLLLWGLYRXVPNRFPARXAFVGALATAFCLETARSLFTWYMGNFDGYRS
    orf144-1     RTAATLTFMTLLLWGLYRFVNRFPARQAFVGALATAFCLETARSLFTWYMGNFDGYRS

55  orf144a.pep  IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRXFDSRGRFDDVLKILLLL
    orf144-1     IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL

    orf144a.pep  DAAQKEGXALPVQEFRRHINMGYDELGELLEKLRHGYIYSGRQGWVLKTGADSIENEL
60  orf144-1     DAAQKEGKALPVQEFRRHINMGYDELGELLEKLRHGYIYSGRQGWVLKTGADSIENEL

    orf144a.pep  FKL FVYRPLPVERDHNQAVDAVMMPCLOT LNMTLAEFDAQAKKQQQS 408
65  orf144-1     FKL FVYRPLPVERDHNQAVDAVMTPCLOT LNMTLAEFDAQAKKRQ 406

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF144 shows 91.2% identity over a 136aa overlap with a predicted ORF (ORF144ng) from *N.gonorrhoeae*:

5	orf144.pep	MTFLQRLQGLADNKICAFWFVRRFDEERVPOXAASMTFTTLLALVPVLTVMVAVASIF	60
	orf144ng	MTFLQCWQGSADNKICAFWFVIRRFSEERVPOAAAASMTFTTLLALVPVLTVMVAVASIF	60
10	orf144.pep	PVFDRWSDSFVSFVNQTIVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID	120
	orf144ng	PVFDRWSDSFVSFVNQTIVPQGADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID	120
	orf144.pep	NTFNRIWRVXXQRPWM	136
15	orf144ng	NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSSGAQQWADAL	180

The complete length ORF144ng nucleotide sequence <SEQ ID 625> is predicted to encode a protein having amino acid sequence <SEQ ID 626>:

	1	MTFLQCWQGS	ADNKICAFW	FVIRRFSEER	VPQAAAASMTF	TTLALVPVL
20	51	TVMVAVASIF	PVFDRWSDSF	VSFVNQTIVP	QGADMVFDYI	DAFRDQANRL
	101	TAIGSVMLV	TSLMLIRTID	NAFNRIWRVN	TQRPWMMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQWADAL	KTAARLAFMT	LLWGLYRFV
	201	PNRFVPARQA	FVGALITAF	LETARFLFTW	YMGNFQGYRS	IYGAFAAVPF
	251	FLLWLNLLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
25	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRQGWVLKT
	351	GADSIELSEL	FKLFVYRPLP	VERDHVNQAV	DAVMTPLQQT	LNMTLAEFDA
	401	QAKKQQQS*				

Further work revealed the following gonococcal DNA sequence <SEQ ID 627>:

	1	ATGACCTTTT	TACAACGTTG	GCAAGGTTTG	GCGGACAATA	AAATCTGTGC
30	51	ATTTGCATGG	TTCGTCATCC	GCCGTTTCAG	TGAAGAGCGC	GTACCGCAGG
	101	CAGCGGCGAG	CATGACGTTT	ACGACACTGC	TGGCACTCGT	CCCCGTACTG
	151	ACCGTAATGG	TCGCGGTCGC	TTCGATTTC	CCCGTGTCG	ACCGCTGGTC
	201	GGATTCTGTC	GTCTCCTTCG	TCAACCAAAC	CATTGTGCCG	CAGGCGCGCG
	251	ATATGGTGTT	CGACTATATC	GACGCATTCC	GCGATCAGGC	AAACCGGCTG
35	301	ACCGCCATCG	GCAGCGTGAT	GCTGGTCGTA	ACCTCGCTGA	TGCTGATTCTG
	351	GACGATAGAC	AATGCGTTCA	ACCGCATCTG	GCGGGTTAAC	ACGCAACGCC
	401	CCTGGATGAT	GCAGTTCCTC	GTTTATTGGG	CGTTGCTGAC	TTTCGGGCCT
	451	TTGTCTTTGG	GTGTGGGCAT	TTCTTTATAG	GTCCGGTCCG	TTCAAGACTC
	501	CGTACTCTCC	TCCGGAGCGC	AACAATGGGC	GGACGCGTTG	AAGACGGCGG
40	551	CAAGCGTGGC	TTTCATGACG	CTTTTGCTGT	GGGGGCTGTA	CCGCTTCGTG
	601	CCCAACCGCT	TCGTGCCCCG	CCGGCAGGCG	TTTGTCCGAG	CTTTGATTAC
	651	GGCATTCTGC	CTGGAGACGG	CACGTTTCCT	GTTCACCTGG	TATATGGGCA
	701	ATTTTCGACG	CTACCGCTCG	ATTTACGGCG	CATTTGCCCG	CGTGCCGTTT
	751	TTCTGCTGT	GGTTAAACCT	GCTGTGGACG	CTGGTCTTGG	GCGGGGCGGT
45	801	GCTGACTTCG	TCGCTGTCTT	ATTGGCAGGG	CGAGGCCTTC	CGCAGGGGAT
	851	TCGACTCGCG	CGGACGGTTT	GACGACGTGT	TGAAAATCCT	GCTGCTTCTG
	901	GATGCGGCGC	AAAAGAAGG	CCGAACCTG	TCCGTTCAGG	AGTTCAGACG
	951	GCATATCAAT	ATGGGTACG	ATGAATTGGG	CGAGCTTTTG	GAAAAGCTGG
	1001	CGCGGTACGG	CTATATCTAT	TCCGGCAGAC	AGGGCTGGGT	TTTGAACACG
50	1051	GGGGCGGATT	CGATTGAGT	GAGCGAATC	TTCAAGCTCT	TCGTGTACCG
	1101	CCCGTTGCct	gtggaAAGGG	ATCATGTGAA	CCAAGCTGtc	gaTGCGGTAA
	1151	TGAcgccgtG	TTTGACAGCT	TTGAACATGA	CGCTGGCGGA	GTTTGACGCT
	1201	CAGgcgAAAA	AACAGCAGCA	GTCTTGA		

This encodes a variant of ORF144ng, having the amino acid sequence <SEQ ID 628; ORF144ng-1>:

55	1	MTFLQRWQGL	ADNKICAFW	FVIRRFSEER	VPQAAAASMTF	TTLALVPVL
	51	TVMVAVASIF	PVFDRWSDSF	VSFVNQTIVP	QGADMVFDYI	DAFRDQANRL
	101	TAIGSVMLV	TSLMLIRTID	NAFNRIWRVN	TQRPWMMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQWADAL	KTAARLAFMT	LLWGLYRFV
	201	PNRFVPARQA	FVGALITAF	LETARFLFTW	YMGNFQGYRS	IYGAFAAVPF
	251	FLLWLNLLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
60	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRQGWVLKT

351 GADSIELSEL FKL FVYRPLP VERDHVNQAV DAVMTPCLQT LNMTLAEFDA  
401 QAKKQQQS\*

ORF144ng-1 and ORF144-1 show 94.1% identity in 406 aa overlap:

```

5      orf144ng-1.pep MTFLQRWQGLADNKICAFWVIRRFSEERVPAQAASMTFTLLALVPVLTVMVAVASIF
      orf144-1      MTFLQRWQGLADNKICAFWVIRRFSEERVPAQAASMTFTLLALVPVLTVMVAVASIF

      orf144ng-1.pep PVFDRWSDSFVSFVNQTI V PQGADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID
10     orf144-1      PVFDRWSDSFVSFVNQTI V PQGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID

      orf144ng-1.pep NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSSGAQQWADAL
      orf144-1      NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDAALASGAPQWGSAL

15     orf144ng-1.pep KTAARLAFMTLLWGLYRFVFNRFVPAQAFV GALITAF CLETARFLFTWYMGNF DGYRS
      orf144-1      RTAATLTFTMTLLWGLYRFVFNRFVPAQAFV GALATAF CLETARSLFTWYMGNF DGYRS

20     orf144ng-1.pep IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL
      orf144-1      IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL

      orf144ng-1.pep DAAQKEGRTLSVQEFRRHINMGYDELGELLEKLARYGYIYSGRQGWVLKTGADSIELSEL
25     orf144-1      DAAQKEGKALPVQEFRRHINMGYDELGELLEKLARHGYIYSGRQGWVLKTGADSIELNEL

      orf144ng-1.pep FKL FVYRPLPVERDHVNQAVDAVMT PCLQTLNMTLAEFDAQAKKQQQS
30     orf144-1      FKL FVYRPLPVERDHVNQAVDAVMT PCLQTLNMTLAEFDAQAKKRQ

```

On this basis of this analysis, including the identification of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 35 Example 75

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 629>:

```

      1 ..AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
      51 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
40    101 GCACCGATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC
      151 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG
      201 CCTGCTTGAA ACACGGGAAC ACGGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 630; ORF146>:

```

      1 ..RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTD MRQE ISALVILLQR
      51 TRRKWLDAHE RQHLRQSLLE TREHG*

```

45 Further work revealed the complete nucleotide sequence <SEQ ID 631>:

```

      1 ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCCTA
      51 CGAACGCTAC CGCTACCGCC GCCTCATCCA CGCCGTCGGG CTCGGCGGGG
100   101 CCGTCCTGTT CGCCACGCCC TCCGCCGGC TGCTCCACCT CCAACACGGC
      151 GAGTGGATAG GGATGACCGT CTTCGTCGTC CTCGGCATGC TCCAGTTTCA
50    201 AGGGGCGATT TACTCCAAGG CGGTGGAACG TATGCTCGGC ACGGTCATCG
      251 GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACCAGCA TTATTTCCAC
      301 GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG
      351 CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCAGGGCTGA
      401 CGATGTGTAT GTCATCGGC GACAACGGCA GCGAATGGCT CGACAGCGGA
55    451 CTCATGCGCG CCATGAACGT CCTCATCGGC GCGGCCATCG CCATCGCCCG

```



-352-

501 CGCCAAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG  
 551 CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC  
 601 AGGCGCATGA CCCGCGAACG CCTCGAGGAG AACATGGCGA AAATGCGCCA  
 651 AATCAACGCA CGCATGGTCA AAAGCCGCGC CCATCTCGCC GCCACATCGG  
 701 GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCCAC  
 751 CGTAAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT  
 801 GCAATCTCCC AAATCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT  
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC  
 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAATGGA  
 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA  
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC  
 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG  
 1101 CCTGCTTGAA ACACGGGAAC ACGGCTGA

This corresponds to the amino acid sequence <SEQ ID 632; ORF146-1>:

15 1 MNTSQNRNLV SRWLNSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG  
 51 EWIGMTVFV LGMLOFQGA YSKAVERMLG TVIGLGAGLG VLWLNQHYFH  
 101 GNLLFYLTVG TASALAGWAA VGKNGYVPM AGLTMCMLIG DNGSEWLD SG  
 151 LMRAMNVLIG AAIAIAAAKL LPLKSTLMWR FMLADNLADC SKMIAEISNG  
 201 RRMTRERLEE NMAKMRQINA RMVKSRSILA ATSGESRISP AMMEAMQHAH  
 251 RKIVNTTELL LTAAKLQSP KLNGSEIRLL DRHFTLLQTD LQQTVALING  
 301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR  
 351 TRRKWLDAHE RQHLRQSLLE TREHG\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF146 shows 98.6% identity over a 74aa overlap with an ORF (ORF146a) from strain A of *N. meningitidis*:

				10	20	30
	orf146.pep			RHARRIRIDTAINPELEALAEHLHYQWQGF		
30	orf146a	KLNGSEIRLLDRHFTLLQTDLQQTVALINGRHARRIRIDTAINPELEALAEHLHYQWQGF				
		280 290 300 310 320 330				
		40 50 60 70				
	orf146.pep	LWLSTDMRQEISALVILLQTRRKWLDAHERQHLRQSLLE TREHGX				
35						
	orf146a	LWLSTNMRQEISALVILLQTRRKWLDAHERQHLRQSLLE TREHSX				
		340 350 360 370				

The complete length ORF146a nucleotide sequence <SEQ ID 633> is:

40 1 ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCCTA  
 51 CGAACGCTAC CGCTACCGCC GCCTCATCCA CGCCGTCCGG CTCGCGGGG  
 101 CCGTCCTGTT CGCCACCGCC TCCGCCCGGC TGCTCCACCT CCAACACGGC  
 151 GAGTGGATAG GGATGACCGT CTTCTGTCGT CTCGGCATGC TCCAGTTTCA  
 201 AGGGGCGATT TACTCCAAGG CCGTGGAACG TATGCTCGGC ACGGTCATCG  
 251 GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACCAGCA TTATTTCCAC  
 45 301 GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG  
 351 CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCGGGGCTGA  
 401 CGATGTGCAT GCTCATCGGC GACAACGGCA GCGAATGGTT CGACAGCGGC  
 451 CTGATGCGCG CGATGAACGT CCTCATCGGC GCGGCCATCG CCATCGCCCG  
 501 CGCCAAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG  
 551 CCGACAACCT GACCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC  
 601 AGGCGCATGA CCCGCGAACG CCTCGAAGAG AACATGGCGA AAATGCGCCA  
 651 AATCAACGCA CGCATGGTCA AAAGCCGCGC CCACCTCGCC GCCACATCGG  
 701 GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCCAC  
 751 CGTAAAATG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT  
 801 GCAATCTCCC AAATCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT  
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC  
 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAATGGA  
 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA  
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC  
 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG  
 1101 CCTGCTTGAA ACACGGGAAC ACAGTTGA

This encodes a protein having amino acid sequence <SEQ ID 634>:

```

      1  MNTSQNRNLV SRWLNSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG
     51  EWIGMTVFV LGMLQFQGA YSKAVERMLG TVIGLGAGLG VLWLNQHYFH
    101  GNLLFYLTVG TASALAGWAA VGKNGYVPM LAGLTMCLIG DNGSEWFDSG
    151  LMRAMNVLIG AAIAIAAAKL LPLKSTLMWR FMLADNLTD CKMIAEISNG
    201  RRMTRERLEE NMAKMRQINA RMVKSRSHLA ATSGESRISP AMMEAMQHAH
    251  RKIVNTTELL LTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTVALING
    301  RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
    351  TRRKWLDAHE RQHLRQSLLE TREHS*

```

10 ORF146a and ORF146-1 show 99.5% identity in 374 aa overlap:

```

      orf146a.pep  MNTSQNRNLVSRWLNSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFV
      orf146-1     MNTSQNRNLVSRWLNSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFV
    15  orf146a.pep  LGMLQFQGAISKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVGTASALAGWAA
      orf146-1     LGMLQFQGAISKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVGTASALAGWAA
    20  orf146a.pep  VGKNGYVPM LAGLTMCLIGDNGSEWFDSGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
      orf146-1     VGKNGYVPM LAGLTMCLIGDNGSEWLD SGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
    25  orf146a.pep  FMLADNLTDCKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
      orf146-1     FMLADNLADCKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
    30  orf146a.pep  AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING
      orf146-1     AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING
    35  orf146a.pep  RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTTRRKWLDAHE
      orf146-1     RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTTRRKWLDAHE
    40  orf146a.pep  RQHLRQSLLE TREHSX
      orf146-1     RQHLRQSLLE TREHG

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

40 ORF146 shows 97.3% identity over a 75aa overlap with a predicted ORF (ORF146ng) from *N.gonorrhoeae*:

```

      orf146.pep  RHARRIRIDTAINPELEALAEHLHYQWQGF 30
      orf146ng    KLNGSEIRLLDRHFTLLQTDLQQTAA LINGRHARRIRIDTAINPELEALAEHLHYQWQGF 364
    45  orf146.pep  LWLSTDMRQEISALVILLQRTTRRKWLDAHERQHLRQSLLE TREHG 75
      orf146ng    LWLSTNMRQEISALVIPLQRTTRRKWLDAHERQHLRQSLLE TREHG 409

```

An ORF146ng nucleotide sequence <SEQ ID 635> was predicted to encode a protein having amino acid sequence <SEQ ID 636>:

```

      1  MSGVRFPSA PIPSTDPPSG SLCFFTFPLQ TASDMNSSQR KRLSGRWLNS
     51  YERYRHRLI HAVRLGGTVL FATALARLLH LQHGEWIGMT VFVVLGMLQF
    101  QGAIYSNAVE RMLGTVIGLG AGLGVWLWNQ HYFHGNLLFY LTIGTASALA
    151  GWAAVGKNGY VPMLAGLTMC MLIGDNGSEW LDSGLMRAMN VLIGAAIAIA
    201  AAKLLPLKST LMWRFMLADN LADCKMIAE ISNGRRMTRE RLEQNMVKMR
    251  QINARMVKSR SHLAATSGES RISPSMMEAM QHAHRKIVNT TELLTTTAAK
    301  LQSPKLNSE IRLDRHFTL LQTDLQQTAA LINGRHARRI RIDTAINPEL
    351  EALAEHLHYQ WQGF LWLSTN MRQEISALVI PLQRTTRKWL DAHERQHLRQ
    401  SLE TREHG*

```

60 Further work revealed the following gonococcal DNA sequence <SEQ ID 637>:

-354-

1 ATGAACTCCT CGCAACGCAA ACGCCTTTCC GgccGCTGGC TCAACTCCTA  
 51 CGAACGCTac cGCCaccGCC GCCTCATACTA TGCCGTGCGG CTCGGCgga  
 101 ccgtCCTGTT CGCCACCGCA CTCGCCCGgc tACTCCACCT CCAacacggc  
 151 gAATGGATAG GGAtgaCCGT CTTCGTCGTC CTCGGCATGC TCCAGTTCCA  
 5 201 AGGCgcgatt tActccaacg cgggtgGAacg taTGctcggg acggtcatcg  
 251 ggctgGGCGC GGGTTTGGgc gTTTATGGC TGAACAGCA TTAtttccac  
 301 ggcaacCTcc tcttctacct gaccatcggc acggcaagcg cactggccgg  
 351 ctGGGCGGCG GTCCGCAAAA acggctacgt ccctatgctg GCGGGGctgA  
 401 CGATGTGCAT gctcatcggc gACAACGGCA GCGAATGGCT CGACAGCGGC  
 10 451 CTGATGCGCG CGATGAACGT CCTCATCGGC GCCGCCATCG CCATTGCCGC  
 501 CGCCAAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG  
 551 CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC  
 601 AGGCGTATGA CGCGCGAACG TTTGGAGCAG AATATGGTCA AAATGCGCCA  
 651 AATCAACGCA CGCATGGTCA AAAGCCGCG CCACCTCGCC GCCACATCGG  
 15 701 GCGAAAGCCG CATCAGCCCC TCCATGATGG AAGCCATGCA GCACGCCCCA  
 751 CGCAAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT  
 801 GCAATCTCCC AAACCTCAACG GCAGCGAAAT CCGGCTGCTC GACCGCCACT  
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGCCGCCCT CATCAACGGC  
 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAAGCTGA  
 20 951 AGCCCTCGCC GACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGTCA  
 1001 GCACCAATAT GCGTCAGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC  
 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG  
 1101 CCTGCTTGAA ACACGGGAAC ACGGCTGA

This corresponds to the amino acid sequence <SEQ ID 638; ORF146ng-1>:

25 1 MNSSQRKRLS GRWLNSYERY RHRRLIHAVR LGGTVLFATA LARLLHLQHG  
 51 EWIGMTVFVV LGMLQFQGA YSNAVERMLG TVIGLGAGLG VLWLNQHYFH  
 101 GNLLFYLTIG TASALAGWAA VGKNGYVPM LAGLTMCLIG DNGSEWLDSG  
 151 LMRAMNVLIG AAIAIAAAKL LPLKSTLMWR FMLADNLADC SKMIAEISNG  
 201 RRMTRERLEQ NMVKMRQINA RMVKSRSHLA ATSGESRISP SMMEAMQHAH  
 30 251 RKIVNTTELL LTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTALING  
 301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR  
 351 TRRKWLDAHE RQHLRQSLLE TREHG\*

ORF146ng-1 and ORF146-1 show 96.5% identity in 375 aa overlap

35 orf146-1.pep MNTSQRNRLVSRWLNSYERYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV  
 orf146ng-1 MNSSQRKRLSGRWLNSYERYRHRRLIHAVRLGGTVLFATALARLLHLQHGEWIGMTVFVV  
 orf146-1.pep LGMLQFQGAIIYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIVGTASALAGWAA  
 40 orf146ng-1 LGMLQFQGAIIYNAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAA  
 orf146-1.pep VGKNGYVPM LAGLTMCLIGDNGSEWLDSGLMRAMNVLIGAAIAIAAAKLPLKSTLMWR  
 orf146ng-1 VGKNGYVPM LAGLTMCLIGDNGSEWLDSGLMRAMNVLIGAAIAIAAAKLPLKSTLMWR  
 45 orf146-1.pep FMLADNLADCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP  
 orf146ng-1 FMLADNLADCSKMIAEISNGRRMTRERLEQNMVKMRQINARMVKSRSHLAATSGESRISP  
 50 orf146-1.pep AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTALING  
 orf146ng-1 SMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTALING  
 orf146-1.pep RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTRRKWLDAHE  
 55 orf146ng-1 RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTRRKWLDAHE  
 orf146-1.pep RQHLRQSLLE TREHG  
 60 orf146ng-1 RQHLRQSLLE TREHG

Furthermore, ORF146ng-1 shows homology with a hypothetical *E.coli* protein:

sp|P33011|YEEA ECOLI HYPOTHETICAL 40.0 KD PROTEIN IN COBU-SBMC INTERGENIC REGION  
 >gi|1736674|gnl|PID|d1016553 (D90838) ORF\_ID:o348#20; similar to [SwissProt  
 Accession Number P33011] [Escherichia coli] >gi|1736682|gnl|PID|d1016560 (D90839)  
 65 ORF\_ID:o348#20; similar to [SwissProt Accession Number P33011] [Escherichia coli]

-355-

>gi|1788318 (AE000292) f352; 100% identical to fragment YEEA\_ECOLI SW: P33011 but has 203 additional C-terminal residues [Escherichia coli] Length = 352  
Score = 109 bits (271), Expect = 2e-23  
Identities = 89/347 (25%), Positives = 150/347 (42%), Gaps = 21/347 (6%)

5 Query: 20 YRHRRLIHAVRLGGTVLFATALARLLHLQHGEWIGMTVFVVLGMLQFOGAIYSNAVERML 79  
YRH R++H R+ L + RL + W +T+ V++G + F G + A ER+  
Sbjct: 15 YRHYRIVHGTRVALAFLLTFLIIRLFTIPESTWPLVTMVMVIMGPISFWGNVVPRAFERIG 74

10 Query: 80 GTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAAVGNKNGYVPMAGLTMCMLI 139  
GTV+G GL L L L + A L GW A+GK Y +L G+T+ +++  
Sbjct: 75 GTVLGSILGLIALQLE---LISLPLMLVWCAAAMFLCGWLALGKKPYQGLLIGVTLAIVV 131

15 Query: 140 GDNGSEWLD SGLMRAMNVLIGXXXXXXXXXXKLPLKSTLMWRFMLADNLADCSKMIAEISN 199  
G E +D+ L R+ +V++G + P ++ + WR LA +L + +++ +  
Sbjct: 132 GSPTGE-IDTALWRS GDVILGSLAMLFETGIWPQRAFIHWRIQLAKSLTEYNRVYQSAFS 190

20 Query: 200 GRRMTRERLEQNMVKMRQINARMVKSRSHLAATSGESRISPSMMEAMQHAHRKIVNXXXX 259  
+ R RLE ++ K+ VK R +A S E+RI S+ E +Q +R +V  
Sbjct: 191 PNLLERPRLESHLQKLL---TDAVKMRGLIAPASKETRI PKSIYEGIQ TINRNLCMLEL 247

25 Query: 260 XXXXXXXXQSPK---LNGSEIRLLDRHFXXXXXXXXXXAALINGRHARRIRIDTAINPEL 316  
+ LN ++R D AL G +N +  
Sbjct: 248 QINAYWATRPSHFVLLNAQKLR--DTQHMMQIILLSLVHALYEGNPQPVFANTEKLNDV 305

Query: 317 EALAEHL--HYQWQ-----GFLWLSTNMQRQEISALVILLQRTTRK 354  
E L + L H+ + G++WL+ ++ L L+ R RK  
Sbjct: 306 EELRQLNNHDLKVVETPIYGYVWLNMETAHQLELLSNLICRALRK 352

On the basis of this analysis, including the identification of several transmembrane domains in the  
gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and  
their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 76

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 639>

35 1 ..GCCGAAGACA CGCGCGTTAC CGCACAGCTT TTGAGCGCGT ACGGCATTCA  
51 GGGCAAATC GTCACTGTGC GCGAACACAA CGAACGGCAG ATGGCGGACA  
101 AGATTGTCGG CTATCTTTCA GACGGCATGG TTGTGGCACA GGTTTCCGAT  
151 GCGGGTACGC CGGCCGTGTG CGACCCGGGC GCGAACTCG CCCGCCCGCT  
201 GCGTGAGGCC GGGTTTAAAG TCGTTCCCGT CGTGGGCGCA AC.GCGGTGA  
251 TGGCGGCTTT GAGCGTGGCC GGTGTGGAAG GATCCGATTT TTATTTCAAC  
40 301 GGTTTGTAC CGCCGAAATC GGGAGAACGC AGGAACTGT TTGCCAAATG  
351 GGTGCGGGCG GCGTTTCTA TCGTCATGTT TGAAACGCCG CACCGCATCG  
401 GTGCAGCGCT TGCCGATATG GCGAACTGT TCCCGAACG CCGATTAATG  
451 CTGGCGCGCG AAATTACGAA AACGTTTGAA ACGTTCTTAA GCGGCACGGT  
501 TGGGGAAT CAGACGGCAT TGTCTGCCGA CGGCGACCAA TCGCGCGGCG  
45 551 AGATGGTGT GGTGCTTTAT CCGCGCAGG ATGAAAAACA CGAAGGCTTG  
601 TCCGAGTCCG CGCAAAACAT CATGAAATC CTCACAGCCG AGCTGCCGAC  
651 CAAACAGGCG GCGGAGCTTG CTGCCAAAT CACGGGCGAG GGAAAGAAAG  
701 CTTTGTACGA T..

This corresponds to the amino acid sequence <SEQ ID 640; ORF147>:

50 1 ..AEDTRVTAQL LSAYGIQKGL VSVREHNERQ MADKIVGYLS DGMVVAQVSD  
51 AGTPAVCDPG AKLARRVREA GFKVVPVGA XAVMAALSVA GVEGSDIFYFN  
101 GFVPPKSGER RKLFAKWVRA AFPIVMFETP HRIGALADM AELFPERRLM  
151 LAREITKTFE TFLSGTVGEI QTALSADGDQ SRGEMVLVLY PAQDEKHEGL  
201 SESAQNIMKI LTAELPTKQA AEAAKITGE GKKALYD..

Further work revealed the complete nucleotide sequence <SEQ ID 641>:

1 ATGTTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC  
51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC  
101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG  
151 CGCGTTACCG CACAGCTTTT GAGCGCGTAC GGCATTCAGG GCAAACCTCGT

-356-

```

201 CAGTGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTGGGCT
251 ATCTTTCAGA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG
301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCCG
351 GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA
401 GCGTGGCCGG TGTGGAAGGA TCCGATTTTT ATTTCAACGG TTTTGTACCG
451 CCGAAATCGG GAGAACGCAG GAAACTGTTT GCCAAATGGG TGCGGGCGCG
501 GTTTCCTATC GTCATGTTTG AAACGCCGCA CCGCATCGGT GCGACGCTTG
551 CCGATATGGC GGAAGTGTTC CCCGAACGCC GATTAATGCT GGCGCGCGAA
601 ATTACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA
651 GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGGCGAG ATGGTGTGCG
701 TGCTTTATCC GGCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG
751 CAAAACATCA TGAAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC
801 GGAGCTTGCT GCCAAATCA CGGCGAGGG AAAGAAAGCT TTGTACGATC
851 TGGCTCTGTC TTGAAAAAC AAATAG

```

15 This corresponds to the amino acid sequence <SEQ ID 642; ORF147-1>:

```

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
51 RVTAQLLSAY GIQKLVSVR EHNERQMAK IVGYLSDGMV VAQVSDAGTP
101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSIVAGVEG SDFYFNGFVP
151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRMLLARE
201 ITKTFETFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA
251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with hypothetical protein ORF286 of *E.coli* (accession number U18997)

ORF147 and *E.coli* ORF286 protein show 36% aa identity in 237aa overlap:

```

25 Orf147: 1 AEDTRVTAQLLSAYGIQKLVSVREHNERQMAKIVGYLSDGMVVAQVSDAGTPAVCDPG 60
    AEDTR T LL +GI +L ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG
    Orf286: 43 AEDTRHTGLLLQHFGINARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTFLINDPG 102

30 Orf147: 61 AKLARRVREXXXXXXXXXXXXXXXXXXXXXEGSDFYFNGFVPPKSGERRKLFKAWVRA 120
    L R RE F + GF+P KS RR
    Orf286: 103 YHLVRTCREAGIRVVPLPGPCAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKAEAE 162

35 Orf147: 121 AFPIVMFETPHRIGALADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALSADGD 179
    ++ +E+ HR+ +L D+ + E R ++LARE+TKT+ET VGE+ + D +
    Orf286: 163 PRTLIFYESTHRLDLSLEDIVAVLGESRYVVLARELTKTWETIHGAPVGELLAWVKEDEN 222

40 Orf147: 180 QSRGEMVLVLYPAQDEKHEGLSESAQNIMKILTAE LPTKQAAELA AKITGEGKKALY 236
    + +GEMVL++ + E L A + +L AELP K+AA LAA+I G K ALY
    Orf286: 223 RRGEMVLIV-EGHKAQEEDLPADALRTLALLQAEPLKKAALAAEIHGVKKNALY 278

```

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF147 shows 96.6% identity over a 237aa overlap with ORF75a from strain A of *N. meningitidis*:

```

45 orf147.pep                                10      20      30
    AEDTRVTAQLLSAYGIQKLVSVREHNERQ
    |||||
orf75a    TLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQKLVSVREHNERQ
           20      30      40      50      60      70

50 orf147.pep                                40      50      60      70      80      90
    MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGAXAVMAALSVA
    |||||
orf75a    MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA
           80      90     100     110     120     130

55 orf147.pep                                100     110     120     130     140     150
    GVEGSDFYFNGFVPPKSGERRKLFKAWVRAAFPIVMFETPHRIGALADMAELFPERRLM
    |||||
orf75a    GVAGSDFYFNGFVPPKSGERRKLFKAWVRAAFPIVMFETPHRIGATLADMAELFPERRLM
           140     150     160     170     180     190

60 orf147.pep                                160     170     180     190     200     210
    LAREITKTFETFLSGTVGEIQTALSADGDQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI

```

```

|||||
orf75a      LAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI
          200      210      220      230      240      250

          220      230
orf147.pep  LTAE LPTKQAAELAAKITGEGKKALYD
          |||||
orf75a      LTAE LPTKQAAELAAKITGEGKKALYDLALS WKNKX
          260      270      280      290

```

### Homology with a predicted ORF from *N.gonorrhoeae*

	orf147.pep	AEDTRVTAQLLSAYGIQGKLVSVREHNERQ           :	30
15	orf147ng	TLYVVATPIGNLADITLRALAVLQADIICAEDTRVTAQLLSAYGIQGRVSVREHNERQ           :	85
	orf147.pep	MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGAXAVMAALSVA           :	90
20	orf147ng	MADKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA           :	145
	orf147.pep	GVEGSDFYFNGFVPPKSGERRKLFAKWVRAAFPIVMFETPHRIGAALADMAELFPERRLM           :	150
	orf147ng	GVAESDFYFNGFVPPKSGERRKLFAKWVRAAFPVVMFETPHRIGATLADMAELFPERRLM           :	205
25	orf147.pep	LAREITKTFETFLSGTVGEIQTALSADGDQSRGEMVLVLVPAQDEKHEGLSESAQNIMKI           :	210
	orf147ng	LAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLVPAQDEKHEGLSESAQNAMKI           :	265
30	orf147.pep	LTAELPTKQAAELAAKITGEGKKALYD   :	237
	orf147ng	LAAELPTKQAAELAAKITGEGKKALYDLALSWKNK   :	300

35	1	MSVFQTAFFM	FQKHLQKASD	SVVGGTLYVV	ATPIGNLADI	TLRALAVLQK
	51	ADIICAEDTP	VTAQLLSAYG	IQGRLVSVRE	HNERQMADKV	IGFLSDGLVV
	101	AQVSDAGTPA	VCDPGAKLAR	RVREAGFKVV	PVVGASAVMA	ALSVAGVAES
	151	DFYFNGFVPP	KSGERRKLFA	KWVRAAFPVV	MFETPHRIGA	TLADMAELFP
	201	ERRLMLAREI	TKTFETFLSG	TVGEIQTALA	ADGNQSRGEM	VLVLYPAQDE
40	251	KHEGLSESAQ	NAMKILAAEL	PTKQAAELAA	KITGEGKKAL	YDLALSWKNN
	301	*				

	1	ATGTTTCAGA	AACACTTGCA	GAAAGCCTCC	GACAGCGTCG	TCGGAGGGAC
45	51	ATTATACGTG	GTTGCCACGC	CCATCGGCAA	TTTGGCAGAC	ATTACCCTGC
	101	GCGCTTTTGC	GGTATTGCAA	AAGGCGGACA	TCATTTGTGC	CGAAGACACG
	151	CGCGTTACTG	CGCAGCTTTT	GAGCGCGTAC	GGCATTCAGG	GCAGGTTGGT
	201	CAGTGTGCGC	GAACACAACG	AGCGGCAGAT	GGCGGACAAG	GTAATCGGTT
	251	TCCTTTCAGA	CGGCCTGGTT	GTGGCGCAGG	TTTCCGATGC	GGGTACGCCG
50	301	GCCGTGTGCG	ACCGGGGCGC	GAAACTCGCC	CGCCGCGTGC	GCGAAGCAGG
	351	GTTCAAAGTC	GTTCCCGTCG	TGGGCGCAAG	CGCGGTAATG	CGCGCGCTGA
	401	GTTGTGCCGG	TGTGGCGGAA	TCCGATTTTT	ATTTCAACGG	TTTGTATCCG
	451	CCGAAATCGG	GCGAACGTAG	GAAATTGTTT	GCCAAATGGG	TGCGGGCGGC
	501	ATTTCTGTGC	GTCATGTTTG	AAACGCGCGA	CCGAATCGGG	GCAACGCTTG
55	551	CCGATATGGC	GGAAATGTTC	CCCGAACGCC	GTCTGATGCT	GGCGCGCGAA
	601	ATCACGAAAA	CGTTTGAAC	GTTCTTAAGC	GGCAGGGTTG	GCGGAAATTCA
	651	GACGGCATTG	GCGCGGACG	GCAACCAATC	GCGCGGCGAG	ATGGTGTGCG
	701	TGCTTTATCC	GGCGCAGGAT	GAAAAACACG	AAGGCTTGTC	CGAGTCTGCG
	751	CAAAATGCGA	TGAAAATCCT	TGCGGCCGAG	CTGCCGACCA	AGCAGGCGGC
60	801	GGAGCTTGCC	GCCAAGATTA	CAGGTGAGGG	CAAAAAGGCT	TTGTACGATT
	851	TGGCACTGTC	TGGGAAAAAC	AAATGA		

This corresponds to the amino acid sequence <SEQ ID 646; ORF147ng-1>:

```

      1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
     51 RVTAQLLSAY GIQGRIVSVR EHNERQMADK VIGFLSDGLV VAQVSDAGTP
    101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVAE SDFYFNGFVP
     5  151 PKSGERRKLF AKWVRAAFPV VMFETPHRIG ATLDMAELF PERRMLLARE
    201 ITKTFETFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
    251 QNAMKILAAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*

```

ORF147ng shows homology to a hypothetical *E.coli* protein:

```

10  sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION
    (F286)
    >gi|606086 (U18997) ORF_f286 [Escherichia coli]
    >gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region
    [Escherichia coli] Length = 286
    Score = 218 bits (550), Expect = 3e-56
    Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)

15  Query: 4   KHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQ 63
      K   Q   A   +S   G   LY+V   TPIGNLADIT   RAL   VLQ   D+I   AEDTR   T   LL   +GI
    Sbjct: 2   KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVLQAVDLIAAEDTRHTGLLLQHFGIN 59

20  Query: 64   GRLVSVREHNERQMADKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123
      RL   ++   +HNE+Q   A+   ++   L   +G   +A   VSDAGTP   +   DPG   L   R   REAG   +VVP+
    Sbjct: 60   ARLFALHDHNEQQKAETLLAKLQEQNIALVSDAGTPLINDPGYHLVRTCREAGIRVVPL 119

25  Query: 124  VGASAVMAALSVAAGVAESDFYFNGFVPPKSGERRKLFKWKVRAAFPVVMFETPHRIGATL 183
      G   A   +   ALS   AG+   F   +   GF+P   KS   RR   ++   +E+   HR+   +L
    Sbjct: 120  PGPCAAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKAEAEPTLIFYESTHRLDLSL 179

30  Query: 184  ADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEK 242
      D+   +   E   R   ++LARE+TKT+ET   VGE+   +   D   N+   +GEMVL++   +
    Sbjct: 180  EDIVAVLGESRYVVLARELTKTWETIHGAPVGEALLAWVKEDENRRKGEMVLIV-EGHKAQ 238

    Query: 243  HEGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKALYDLAL 286
      E   L   A   +   +L   AELP   K+AA   LAA+I   G   K   ALY   AL
35  Sbjct: 239  EEDLPADALRTLALLQAEPLKKAALAAEIHGVKKNALYKYAL 282

```

Based on the computer analysis and the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 77

40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 647>

```

      1 ATGAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA
     51 AACCGGTCGC ATCCGCTTCT C.GCTGCTTA CTTAGCCATA TGCCTGTCGT
    101 TCGGCATTCT TCCCCAAGCC TGGGCGGGAC ACACTTATTT CGGCATCAAC
     5  151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
    201 GGCGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGCGAAAT
    251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAACGGC
    301 GTGGCGGCAT TGGTGGGCGT ATCAATATAT TGTGAGCGTG GCACATAACG
    351 GCGGCTATAA CAACGTTGAT TTTGGTGCGG AAGGAak.AA tATCCC.GAT
    401 CAACAwCGww TTACTTATAA AATTGTGAAA CGGAATAATT ATAAAGCAGG
     5  451 GACTAAAGGC CATCCTTAGC GCGGCGATTA TCATATGCCG CGTTTGATA
    501 AATwTGTCAC AGATGCAGAA CCTGTTGAAA TGACCAGTTA TATGGATGGG
    551 CGGAAATATA TCGATCAAAA TAATTACCCT GACCGTGTTT GTATTGGGGC
    601 AGGCAGGCAA TATTGGCGAT CTGATGAAGA TGAGCCCAAT AACCGCGAAA
    651 GTTCATATCA TATTGCAAGT .....
    701 ..... GGCTC ACCAATGTTT ATCTATGATG CCCAAAAGCA
    751 AAAGTGGTTA ATTAATGGGG TATTGCAAAAC GGGCAACCCC TATATAGGAA
    801 AAAGCAATGG CTTCCAGCTG GTTCGTAAAG ATTGGTTCTA TGATGAAATC
    851 TTTGCTGGAG ATACCCATTC AGTATTCTAC GAACCCAGTC AAAATGGGAA
    901 ATACTCTTTT AACGACGATA ATAATGGCAC AGGAAAAATC AATGCCAAAC

```

-359-

951 ATGAACACAA TTCTCTGCCT AATAGATTAA AAACACGAAC CGTTCAATTG  
 1001 TTTAATGTTT CTTTATCCGA GACAGCAAGA GAACCTGTTT ATCATGCTGC  
 1051 AGGTGGTGTC AACAGTTATC GACCCAGACT GAATAATGGA GAAAATATTT  
 1101 CCTTTATTGA CGAAGGAAAA GGCGAATTGA TACTTACCAG CAACATCAAT  
 1151 CAAGGTGCTG GAGGATTATA TTTCCAAGGA GATTTTACGG TCTCGCTGA  
 1201 AAATAACGAA ACTTGGCAAG GCGCGGGCGT TCATATCAGT GAAGACAGTA  
 1251 CCGTTACTTG GAAAGTAAAC GCGGTGGCAA ACGACCGCCT GTCCAAAATC  
 1301 GGCAAAGGCA CGCTG.....  
 //  
 2101 ..... GATAAAG  
 2151 TGACTGCTTC ATTGACTAAG ACCGACATCA GCGGCAATGT CGATCTTGCC  
 2201 GATCACGCTC ATTTAAATCT CACAGGGCTT GCCCACTCA ACGGCAATCT  
 2251 TAGTGCAAAT GGCGATACAC GTTATACAGT CAGCCACAAC GCCACCCAAA  
 2301 ACGGCAACCK TA<sub>g</sub>CctCGtG G.sAATGcCC AAGCAACATT TAATCAAGCC  
 2351 ACATTAAACG GCAACACATC GGCTTCgGGC AATGCTTCAT TTAATCTAAG  
 2401 CGACCACGCC GTACAAAACG GCAGTCTGAC GCTTTCCGGC AACGCTAAGG  
 2451 CAAACGTAAG CCATTCCGCA CTCAACGGTA ATGTCTCCCT AGCCGATAAG  
 2501 GCAGTATTCC ATTTTGAAAG CAGCCGCTTT ACCGGACAAA TCAGCGGGCG  
 2551 Ca<sub>g</sub>GATACG GCATTACACT TAAAAGACAG CGAATGGACG CTGCCGTCA<sub>g</sub>  
 2601 GarCGGAATT AGGCAATTTA AACCTTGACA ACGCCACCAT TACaCTCAAT  
 2651 TCCGCCTATC GCCACGATGC GGCAGGGGCG CAAACCGGCA GTGCGACAGA  
 2701 TGCGCCGCGC CGCCGTTCGC GCCGTTCGCG CCGTTCCTTA TTATmCGTTA  
 2751 CACCGCCAAC TTCGGTAGAA TCCCGTTTCA ACACGCTGAC GGTAAACGGC  
 2801 AAATTGAACG GTACGGGAAC ATTCCGCTTT ATGTCGGAAC TCTTCGGCTA  
 2851 CCGCAGCGAC AATTTGAAGC TGGCGGAAAG TTCCGAAGGC ACTTACACCT  
 2901 TGGCGGTCAA CAATACCGGC AACGAACCTG CAAGCCTCGA ACAATTGACG  
 2951 GTAGTGAAG GAAAAGACAA CAAACCGCTG TCCGAAAACC TTAATTTAC  
 3001 CCTGCAAAAC GAACACGTCG ATGCAGGCGC GTG.....  
 //  
 3551 ..... TTAGAC CGCGTATTTG CCGAAGACCG  
 3601 CCGCAACGCC GTTTGGACAA GCGGCATCCG GGACACCAA CACTACCGTT  
 3651 CGCAAGATTT CCGCGCCTAC CGCCAACAAA CCGACCTGCG CCAAATCGGT  
 3701 ATGCAGAAAA ACCTCGGCAG CGGGCGCGTC GGCACTCCTGT TTTCGCACAA  
 3751 CCGGACCGAA AACACCTTCG ACGACGGCAT CGGCAACTCG GCACGGCTTG  
 3801 CCCACGGCGC CGTTTTCGGG CAATACGGCA TCGACAGGTT CTACATCGGC  
 3851 ATCAG<sub>n</sub>CGCG GGCGCGGGTT TTAGCAGCGG CAGCCTTTcA GACGGCATCG  
 3901 GAGsmAAAwT CCGCCGCGCG GTGctGCATT ACGGCATTCA GGCACGAtAC  
 3951 CGCGCCGgtt tCg<sub>g</sub>CGgAtT CGGCATCGAA CCGCACATCG GCGCAACGCG  
 4001 ctATTTCTGTC CAAAAGCGG ATTACCGCTA CGAAAACGTC AATATCGCCA  
 4051 CCCCCGGCCT TGCATTCAAC CGcTACCGCG CGGGCATTa<sub>a</sub> GGCAGATTAT  
 4101 TCATTCAAAC CGGCGCAACA CATTTCATC ACGCCTTATT TGAGCCTGTC  
 4151 CTATACCGAT GCCGCTTCGG GCAAAGTCG AACACGCGTC AATACCGCCG  
 4201 TATTGGCTCA GGATTTCCGG AAAACCCGCA GTGCGGAATG GGgCGTAAAC  
 4251 GCCGAATCA AAGGTTTCAC GCTGTCCCTC CACGCTGCCG CCGCCAAAGG  
 4301 CCCGCAACTG GAAGCGCAAC ACAGCGCGGG CATCAAATTA GGCTACCGCT  
 4351 GGTAA...

This corresponds to the amino acid sequence <SEQ ID 648; ORF1>:

1 MKTTDKRTE THRKAPKTGR IRFXAAYLAI CLSFGILPQA WAGHTYFGIN  
 51 YQYYRDFAEK KGKFAVGAKE IEVYNKKGEL VGKSMTKAPM IDFSVSVSRNG  
 101 VAALVGVQYI VSVAHNGGYN NVDFGAEGXN IXDQXRXTYK IVKRNNYKAG  
 151 TKGHFYGDDY HMPRLHKXVT DAEPVEMTSY MDGRKYIDQN NYPDRVRIGA  
 201 GRQYWRSDDED EPNNRESSYH IAS.....GS PMFIYDAQKQ  
 251 KWLINGVLQT GNPYIGKSNG FQLVRKDWFY DEIFAGDTHS VFYEPRQNGK  
 301 YSFNDNNGT GKINAKHEHN SLPNRLKTRT VQLEFVSLSE TAREPVYHAA  
 351 GGVNSYRPRL NNGENISFID EGKGELILTS NINQGAGGLY FQGDFTVSPE  
 401 NNETWQAGAV HISEDSTVTW KLVGVANDRL SKIGKGTLL.....  
 //  
 701 .....DKVTAS LTKTDISGNV DLADHAHLNL TGLATLNGNL  
 751 SANGDTRYTV SHNATQNGNX SLVXNAQATF NQATLNGNTS ASGNASFNLS  
 801 DHAVQNGSLT LSGNAKANVS HSA LNNGVSL ADKAVFHFES SRFTGQISGG  
 851 KDTALHLKDS EWTLP SGXEL GNLNL DNATI TLNSAYRHDA AGAQTGSATD  
 901 APRRRSRRSR RSL LXVT PPT SVESRENTLT VNGKLNQGT FRFMSELFY  
 951 RSDKLKLAES SEGTYTLAVN NTGNEPASLE QLTVEGKDN KPLSENLFNT  
 1001 LQNEHVDAGA W.....  
 //  
 1151 .....LDRVFAEDR  
 1201 RNAVWTSGIR DTKHYRSQDF RAYRQQTDLR QIGMQKNLGS GRVGILFSDH  
 1251 RENTENTFDDGI GNSARLAHGA VFGQY GIDRF YIGISAGAGF SSGSLSDGIG  
 1301 XKXRRRVLHY GIQARYRAGF GGFGEI PHIG ATRYFVQKAD YRYENVNIAT  
 1351 PGLAFNRYRA GIKADYSFKP AQHISITPYL SLSYTDASG KVRTRVNTAV



-360-

1401 LAQDFGKTRS AEWGVNAEIK GFTLSLHAAA AKGPQLEAQH SAGIKLGYRW  
 1451 \*

Further sequencing analysis revealed the complete nucleotide sequence <SEQ ID 649>:

5	1	ATGAAAACAA	CCGACAAACG	GACAACCGAA	ACACACCGCA	AAGCCCCGAA
	51	AACCGGCCGC	ATCCGCTTCT	CGCCTGCTTA	CTTAGCCATA	TGCCTGTCTG
	101	TCGGCATTCT	TCCCCAAGCC	TGGGCGGGAC	ACACTTATTT	CGGCATCAAC
	151	TACCAATACT	ATCGCGACTT	TGCCGAAAAT	AAAGGCAAGT	TTGCAGTCGG
	201	GGCGAAAGAT	ATTGAGGTTT	ACAACAAAAA	AGGGGAGTTG	GTCCGGCAAT
10	251	CAATGACAAA	AGCCCCGATG	ATTGATTTTT	CTGTGGTGTC	GCGTAACGGC
	301	GTGGCGGCAT	TGGTGGGCGA	TCAATATATT	GTGAGCGTGG	CACATAACGG
	351	CGGCTATAAC	AACGTTGATT	TTGGTGCAGA	AGGAAGAAAT	CCCGATCAAC
	401	ATCGTTTTAC	TTATAAAATT	GTGAAACGGA	ATAATTATAA	AGCAGGGACT
	451	AAAGGCCATC	CTTATGGCGG	CGATTATCAT	ATGCCGCGTT	TGCATAAATT
	501	TGTCACAGAT	GCAGAACCTG	TTGAAATGAC	CAGTTATATG	GATGGGCGGA
15	551	AATATATCGA	TCAAAATAAT	TACCCTGACC	GTGTTCGTAT	TGGGGCAGGC
	601	AGGCAATATT	GGCGATCTGA	TGAAGATGAG	CCCAATAACC	GCGAAAGTTC
	651	ATATCATATT	GCAAGTTCGT	ATTCTTGGCT	CGTTGGTGGC	AATACCTTTG
	701	CACAAAATGG	ATCAGGTGGT	GGCACAGTCA	ACTTAGGTAG	TGAAAAAATT
	751	AAACATAGCC	CATATGGTTT	TTTACCAACA	GGAGGCTCAT	TTGGCGCAGC
20	801	TGGCTCACCA	ATGTTTATCT	ATGATGCCCA	AAAGCAAAAG	TGGTTAATTA
	851	ATGGGGTATT	GCAAACGGGC	AACCCCTATA	TAGGAAAAAG	CAATGGCTTC
	901	CAGCTGGTTC	GTAAGGATTG	GTTCTATGAT	GAAATCTTTG	CTGGAGATAC
	951	CCATTCACTA	TTCTACGAAC	CACGTCAAAA	TGGGAAATAC	TCTTTTAACG
	1001	ACGATAATAA	TGGCAGCAGA	AAAATCAATG	CCAAACATGA	ACACAATTCT
25	1051	CTGCCTAATA	GATTAAAAAC	ACGAACCGTT	CAATTGTTTA	ATGTTTCTTT
	1101	ATCCGAGACA	GCAAGAGAAC	CTGTTTATCA	TGCTGCAGGT	GGTGTCAACA
	1151	GTTATCGACC	CAGACTGAAT	AATGAGAGAA	ATATTTCTCT	TATTGACGAA
	1201	GGAAAAGGCG	AATTGATACT	TACCAGCAAC	ATCAATCAAG	GTGCTGGAGG
	1251	ATTATATTTC	CAAGGAGATT	TTACGGTCTC	GCCTGAAAAT	AACGAAACTT
30	1301	GGCAAGGCGC	GGGCGTTCAT	ATCAGTGAAG	ACAGTACCGT	TACTTGGAAG
	1351	GTAACGCGCG	TGGCAAACGA	CCGCTGTCTC	AAAATCGGCA	AAGGCACGCT
	1401	GCACGTTCAA	GCCAAAGGGG	AAAACCAAGG	CTCGATCAGC	GTGGGCGACG
	1451	GTACAGTCAT	TTTGGATCAG	CAGGCAGACG	ATAAAGGCAA	AAAACAAGCC
	1501	TTTAGTGAAA	TCGGCTTGGT	CAGCGGCAGG	GGTACGGTGC	AACTGAATGC
35	1551	CGATAATCAG	TTCAACCCCG	ACAACTCTA	TTTCGGCTTT	CGCGGCGGAC
	1601	GTTTGGATTT	AAACGGGCAT	TCGCTTTCGT	TCCACCGTAT	TCAAAATACC
	1651	GATGAAGGGG	CGATGATTGT	CAACCACAAT	CAAGACAAAG	AATCCACCGT
	1701	TACCATTACA	GGCAATAAAG	ATATTGCTAC	AACCGGCAAT	AACAACAGCT
40	1751	TGGATAGCAA	AAAAGAAATT	GCCTACAACG	GTGGTTTGGG	CGAGAAAGAT
	1801	ACGACCAAAA	CGAACGGGCG	GCTCAACCTT	GTTTACCAGC	CCGCCGAGA
	1851	AGACCGCACC	CTGCTGCTTT	CCGGCGGAAC	AAATTTAAAC	GGCAACATCA
	1901	CGCAACAAAA	CGGCAAACTG	TTTTTCAGCG	GCAGACCAAC	ACCGCACGCC
	1951	TACAATCATT	TAAACGACCA	TTGGTCGCAA	AAAGAGGGCA	TTCTCGCGG
	2001	GGAATCTGTG	TGGGACAACG	ACTGGATCAA	CCGCACATT	AAAGCGGAAA
45	2051	ACTTCCAAAT	TAAAGCGCGA	CAGGCGGTGG	TTTCCCGCAA	TGTTGCCAAA
	2101	GTGAAAGGCG	ATTGGCATT	GAGCAATCAC	GCCCAAGCAG	TTTTTGGTGT
	2151	CGCACCAGAT	CAAAGCCACA	CAATCTGTAC	ACGTTCCGAC	TGGACGGGTC
	2201	TGCAAAATTG	TGTCGAAAAA	ACCATTACCG	ACGATAAAGT	GATTGCTTCA
	2251	TTGACTAAGA	CCGACTCAG	CGGCAATGTC	GATCTTGCCG	ATCAGCTCA
50	2301	TTTAAATCTC	ACAGGGCTTG	CCCACTCAA	CGGCAATCTT	AGTGCAATG
	2351	GCGATACACG	TTATACAGTC	AGCCACAACG	CCACCCAAAA	CGGCAACCTT
	2401	AGCCTCGTGG	GCAATGCCCA	AGCAACATTT	AATCAAGCCA	CATTAAACGG
	2451	CAACACATCG	GCTTCGGGCA	ATGCTTCATT	TAATCTAAGC	GACCACGCCG
	2501	TACAAAACGG	CAGTCTGACG	CTTTCGGGCA	ACGCTAAGGC	AAACGTAAGC
55	2551	CATTCCGCAC	TCAACGGTAA	TGTCTCCCTA	GCCGATAAGG	CAGTATTCCA
	2601	TTTTGAAAGC	AGCCGCTTTA	CCGACAAAAT	CAGCGGCGGC	AAGGATACGG
	2651	CATTACACTT	AAAAGACAGC	GAATGGACGC	TGCCGTCAGG	CACGGAATTA
	2701	GGCAATTTAA	ACCTTGACAA	CGCCACCATT	AACTCAATT	CCGCCTATCG
	2751	CCACGATGCG	CGAGGGGCGC	AAACCGGCAG	TGCGACAGAT	GCGCCGCGCC
60	2801	GCGGTTTCGCG	CCGTTTCGCG	CGTTCCCTAT	TATCCGTTAC	ACCGCCAAT
	2851	TCGGTAGAAT	CCCGTTTCAA	CACGCTGACG	GTAACCGGCA	AATTGAACGG
	2901	TCAGGGAACA	TTCCGCTTTA	TGTCGGAAC	CTTCGGCTAC	CGCAGCGACA
	2951	AATTGAAGCT	GGCGGAAAGT	TCCGAAGGCA	CTTACACCTT	GGCGGTCAAC
	3001	AATACCGGCA	ACGAACCTGC	AAGCCTCGAA	CAATTGACGG	TAGTGGAAGG
65	3051	AAAAGACAAC	AAACCGCTGT	CCGAAAACCT	TAATTTACAC	CTGCAAAACG
	3101	AACACGTCGA	TGCCGGGCGG	TGGCGTTACC	AACTCATCCG	CAAAGACGGC
	3151	GAGTTCCGCC	TGCATAATCC	GGTCAAAGAA	CAAGAGCTTT	CCGACAAACT
	3201	CGGCAAGGCA	GAAGCCAAAA	AACAGGCGGA	AAAAGACAAC	GCGCAAAGCC
	3251	TTGACGCGCT	GATTTCGGCC	GGGCGCGATG	CCGTGCAAAA	GACAGAAAGC
70	3301	GTTGCCGAAC	CGGCCCGGCA	GGCAGGCGGG	GAAAATGTCT	GCATTATGCA

-361-

3351 GCGGAGGAA GAGAAAAAAC GGGTGCAGGC GGATAAAGAC ACCGCTTGG  
 3401 CGAAACAGCG CGAAGCGGAA ACCCGGCCGG CTACCACCGC CTCCCCCGC  
 3451 GCCCGCCGCG CCCGCCGGA TTTGCCGCAA CTGCAACCCC AACCGCAGCC  
 5 3501 CCAACCGCAG CGCGACCTGA TCAGCCGTTA TGCCAATAGC GGTTCAGTG  
 3551 AATTTTCCGC CACGCTCAAC AGCGTTTCG CCGTACAGGA CGAATTAGAC  
 3601 CGCGTATTG CCGAAGACCG CCGCAACGCC GTTTGGACAA GCGGCATCCG  
 3651 GGACACCAA CACTACCGTT CGCAAGATT CCGCGCCTAC CGCCAACAAA  
 3701 CCGACCTGCG CCAAATCGGT ATGCAGAAAA ACCTCGGCAG CGGCGCGTC  
 3751 GGCATCCTGT TTTCCGACAA CCGGACCGAA AACACCTTCG ACGACGGCAT  
 10 3801 CGGCAACTCG GCACGGCTTG CCCACGGCGC CGTTTTCGGG CAATACGGCA  
 3851 TCGACAGGTT CTACATCGGC ATCAGCGCGG GCGCGGGTTT TAGCAGCGGC  
 3901 AGCCTTTCAG ACGGCATCGG AGGCAAAATC CGCCGCCGCG TGCTGCATTA  
 3951 CGGCATTAG GCACGATACC GCGCCGGTTT CGGCGGATTG GGCATCGAAC  
 4001 CGCACATCGG CGCAACGCGC TATTTCTGTC AAAAAGCGGA TTACCGCTAC  
 15 4051 GAAAACGTCA ATATCGCCAC CCCCGCCCTT GCATTCAACC GCTACCGCGC  
 4101 GGGCATTAG CGAGATTATT CATTCAAACC GGCGCAACAC ATTTCCATCA  
 4151 CGCCTTATT GAGCCTGTCC TATACCGATG CCGCTTCGGG CAAAGTCCGA  
 4201 ACACGCGTCA ATACCGCCGT ATTGGCTCAG GATTTCGGCA AAACCCGCGC  
 4251 TGCGGAATGG GCGGTAAACG CCGAAATCAA AGGTTTCACG CTGTCCCTCC  
 20 4301 ACGCTGCCG CGCCAAAGC CCGCAACTGG AAGCGCAACA CAGCGCGGGC  
 4351 ATCAAATTAG GCTACCGCTG GTAA

This corresponds to the amino acid sequence <SEQ ID 650; ORF1-1>:

1 MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN  
 25 51 YQYYRDFAE NKGKFAVGA KD IEVYNKKGEL VGKSMKAPM IDFSVVS RNG  
 101 VAALVGDQYI VSAHNGGYN NVDFGAEGRN PDQHRFTYKI VKRNNYKAGT  
 151 KGHYPYGGDYH MPRHLKFTVD AEPVEMTSYM DGRKYIDQNN YPDRVRIGAG  
 201 RQYWRSEDE PNNRESSYHI ASAYSWLVGG NTFAQNGSGG GTVNLGSEKI  
 251 KHSPYGFLEPT GGSFGDSGSP MFIYDAQKQK WLVINGVLQTG NPYIGKSNFG  
 301 QLVKRDWFDY EIFAGDTHSV FYEPRQNGKY SFNDDNNGTG KINAKHEHNS  
 351 LPNRLKTRTV QLFNVSLSET AREPVYHAAG GVNSYRPRLN NGENISFIDE  
 401 KGELILTSN INQAGGLYF QGDFTVSPEN NETWQAGVH ISEDSTVTWK  
 451 VNGVANDRLS KIGKGLHVG AKGENQGSIS VGDGTVILDQ QADDKGGKQA  
 501 FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSFHRIQNT  
 551 DEGAMIVNHN QDKESTVTIT GNKDIATTGN NNSLDSKKEI AYNGWFGEKD  
 35 601 TTKTNGRLNL VYQPAEDRT LLLSGGTNLN GNITQTNGKL FFSGRPTPHA  
 651 YNHLNDHWSQ KEGIPRGEIV WDNDWINRTF KAENFQIKGG QAVVSRNVAK  
 701 VKGDWHLNSH AQAVFGVAPH QSHITCTRSD WTGLTNCVEK TITDDKVIAS  
 751 LTKTDISGNV DLADHAHLNL TGLATLNGNL SANGDTRYTV SHNATQNGNL  
 801 SLVGNAQATF NQATLNGNTS ASGNASEFNL DHAVQNGSLT LSGNAKANVS  
 40 851 HSA LNVSL ADKAVFHFES SRFTGQISGG KDTALHLKDS EWTLPSTEL  
 901 GNLNLNATI TLNSAYRHDA AGAQTGSATD APRRRSRRSR RSLLSVTPPT  
 951 SVESRFNTLT VNGKLNGQGT FRFMSELFY RSDKLKLAES SEGTYTLAVN  
 1001 NTGNEPASLE QLTVEGKDN KPLSENLFNT LQNEHVDAGA WRYQLIRKDG  
 1051 EFRHLNPVKE QELSDKLGA EAKKQAEKDN AQSLDALIAA GRDAVEKTES  
 45 1101 VAEPARQAGG ENVGIMQAE EKRVQADKD TALAKQREAE TRPATTAFFR  
 1151 ARRARDLPQ LQPQPQPQ RDLISRYANS GLSEFSATLN SVFAVQDELD  
 1201 RVFAEDRRNA VWTSGIRDIT HYRSQDFRAY RQQTDLRQIG MQKNLGSGRV  
 1251 GILFSHNRT NTFDDGIGNS ARLAHGAVFG QYGIDRFYIG ISAGAGFSSG  
 1301 SLSDGIGGKI RRRVLHYGIQ ARYRAGFGG GIEPHIGATR YFVQKADYRY  
 50 1351 ENVNIATPGL AFNRYRAGIK ADYSFKPAQH ISITPYLSLS YTDAASGKVR  
 1401 TRVNTAVLAQ DFGKTRSAEW GVNAEIKGFT LSLHAAAAGK PQLEAQHSAG  
 1451 IKLGYRW\*

Computer analysis of these sequences gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

55 ORF1 shows 57.8% identity over a 1456aa overlap with an ORF (ORF1a) from strain A of *N. meningitidis*:

60 orf1.pep MKTTDKRTTETHR KAPKTGRIRFXAAYLAICLSFGILPQAWAGHTYFGIN YQYYRDFAE N  
 orf1a MKTTDKRTTETHR KAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGIN YQYYRDFAE N  
 orf1.pep KGKFAVGA KD IEVYNKKGEL VGKSMKAPM IDFSVVS RNGVAALVGVQYIVSAHNGGYN

[illegible]

70 The complete length ORF1a nucleotide sequence <SEQ ID 651> is:

	1	ATGAAAACAA	CCGACAAACG	GACAACCGAA	ACACACCGCA	AAGCCCCGAA
	51	AACCGGCCGC	ATCCGCTTCT	CGCCTGCTTA	CTTAGCCATA	TGCCTGTCTG
	101	TCGGCATTCT	TCCCAAGCT	TGGGCGGGAC	ACACTTATTT	CGGCATCAAC
5	151	TACCAATACT	ATCGCGACTT	TGCCGAAAAT	AAAGGCAAGT	TTGCAGTCGG
	201	GGCGAAAGAT	ATTGAGTNT	ACAACAAAAA	AGGGGAGTTG	GTCCGCAAAAT
	251	CAATGACAAA	AGCCCCGATG	ATTGATTTTT	CTGTGGTGTC	GCCTAACGGC
	301	GTGGCGGCAT	TGGTGGGCGA	TCAATATATT	GTGAGCGTGG	CACATAACGG
	351	CGGCTATAAC	AACGTTGATT	TTGGTGCGGA	AGGAAGNAAT	CCCGATCAGC
	401	ACCGTTTTTC	TTACCAAATT	GTGAAAAGAA	ATAATTATAA	GCCTGACAAT
10	451	TCACACCCCT	ACAACGGCGA	TTANCATATG	CCGCGTTTGC	ATAAATTTGT
	501	CACAGATGCA	GAACCTGTCT	AAATGACGAG	TGACATGAGG	GGGAATACCT
	551	ATTCCGATAA	AGAAAAATAT	CCCGAGCGTG	TCCGCATCGG	CTCAGGACAC
	601	CACTATTGGC	GTTATGATGA	TGACAAACAC	GGCGATTTAT	CCTACTCCGG
	651	CGCATGGTTA	ATTGGCGGCA	ATACACATAT	GCAGGGTTGG	GGAAATAATG
15	701	GCGTANTTAG	TTTGAGCGGC	GATGTGCGCC	ATGCCAACGA	CTATGGCCCT
	751	ATGCCGATTG	CAGGTGCGGC	AGGCGACAGC	GGTTCGCCAA	TGTTTATTTA
	801	TGACAAAACA	AACAATAAAT	GGCTGCTCAA	CGGAGTTTTA	CAAACCGGCT
	851	ACCCTTATTC	CGGCAGGGAA	AACGTTTCC	AGCTGATACG	CAAAGATTGG
	901	TTCTACGATG	ACATTTACAG	AGGCGATACA	CATACCGTCT	NTTTGAACC
20	951	GGCGAGTAAC	GGACATTTT	CCTTTACATC	CAACAACAAC	GGTACGGGTA
	1001	CGGTAACAGA	AACCAACGAA	AAGGTNTCCA	ATCCAAAGCT	TAAAGTACAG
	1051	ACAGTCCGAC	TGTTTGACGA	ATCTTTGAAT	GAAACTGATA	AAGAACCAGT
	1101	TTACGCGGCA	GGGGGTGTTA	ATCAGTACCG	TCCAAGGTTA	AACAACGGTG
	1151	AAAACCTTTC	TTTTATCGAT	TACGGCAACG	GCAAACCTCAT	CTTATCAAAC
25	1201	AACATCAACC	AAGGCGGGG	CGGTTGTAT	TTTGAAGGTG	ATTTTACGGT
	1251	CTCGCCTGAA	AACAACGAAA	CGTGCCAAGG	CGCGGGCGTT	CATATCAGTG
	1301	AAGACAGTAC	CGTTACTTGG	AAAGTAAACG	GCGTGGCAAA	CGACCGCCTG
	1351	TCCAAAATCG	GCAAAGGCAC	GCTGCACGTT	CAAGCCAAAG	GGGAAAACCA
	1401	AGGCTCGATC	AGCGTGGGCG	ACGGTACAGT	CATTTTGGAT	CAGCAGGCAG
30	1451	ACGATAAAGG	CAAAAAACAA	GCCTTAGTGT	AAATCGGCTT	GNTCAGCGGC
	1501	AGGGGTACGG	TGCAACTGAA	TGCCGATAAT	CAGTTCAACC	CCGACAAACT
	1551	CTATTTCCGG	TTTCGCGGCG	GACGTTTGGA	TTTAAACGGG	CATTCGCTTT
	1601	CGTTCCACCG	TATTCAAAT	ACCGATGAAG	GGGCGATGAT	TGNCNATCAT
	1651	AATGCCACAA	CAACATCCAC	CGTTACCATT	ACAGGGAATG	AAAGTATTAC
35	1701	ACAACCGAGT	CGTAAGAATA	TCAATAGACT	TAATTACAGC	AAAGAAATTG
	1751	CCTACAACGG	TTGGTTTGGC	GAGAAAGATA	CGACCAAAAC	GAACGGGCGG
	1801	CTCAACCTTG	TTTACCAGCC	CGCCGAGAA	GACCGCACCC	NGCTGCTTTC
	1851	CGGCGGAACA	AATTTAAACG	GCAACATCAC	GCAAACAAAC	GGCAAACCTG
	1901	TTTTCAGCGG	CAGACCGACA	CCGCACGCCT	ACAATCATT	AGGAAGCGGG
40	1951	TGGTCAAAA	TGGAAGGTAT	CCCACAAGGA	GAAATCGTGT	GGGACAACGA
	2001	CTGGATCNAC	CGCACGTTTA	AAGCGGAAAA	TTTCCATATT	CAGGGCGGGC
	2051	AGGCGGTGAT	TTCCCGCAAT	GTTGCCAAAG	TGGAAGGCGA	TTGNCATTTG
	2101	AGCAATCACG	CCCAAGCAGT	TTTTGGTGTC	GCACCGCATC	AAAGCCATAC
	2151	AATCTGTACA	CGTTCGGACT	GGACNGGTCT	GACAAATTGT	GTGCAANAAA
45	2201	NCATTACCGA	CGATAAAGTG	ATTGCTTCAT	TGACTAAGAC	NGACNTNAGC
	2251	GGCANTGTNA	GNCTNNCCNA	TNACGNTNNT	TNAAANCTCN	CNGGGCNTGC
	2301	NNCACTNAAN	GGCAATCTTA	GTGCAAATGG	CGATACACGT	TATACAGTCA
	2351	GCCACAACGC	CACCCAAAAC	GGCAACCTTA	GCCTCGTGGG	CAATGCCCAA
	2401	GCAACATTTA	ATCAAGCCAC	ATTAAACGGC	AACNCATCGG	NTTCGGGCAA
50	2451	TGCTTCATTT	AATCTAAGCA	ACAACGCCGC	ACAAAACGGC	AGTCTGACGC
	2501	TTTCCGACAA	CGCTAAGGCA	AACGTAAGCC	ATTCCGCACT	CAACGGCAAT
	2551	GTCTCCCTAG	CCGATAAGGC	AGTATTCCAT	TTTGAACACA	GCCGCTTTAC
	2601	CGGACAATC	AGCGGCAGCA	AGGANACAGC	ATTACACTTA	AAAGACAGCG
	2651	AATGGACGCT	GCCGTACGGC	ACGGAATTAG	GCAATTTAAA	CCTTGACAAC
55	2701	GCCACCATTA	CACCTAATTC	CGCCTATCGC	CACGATGCTG	CAGGCGCGCA
	2751	AACCGGCAGN	GTGTCAGACA	CGCCGCGCCG	CCGTTCGCGC	CGTTCCCTAT
	2801	TATCCGTTAC	ACCGCCAAT	TCGGTAGAAT	CCCGTTTCAA	CACGCTGACG
	2851	GTAAACGGCA	AATTGAACNG	TCAAGGAACA	TTCCGCTTTA	TGTCGGAACT
	2901	CTTCGGCTAC	CGAAGCGACA	AATTGAAGCT	GGCGGAAAGT	TCCGAAGGNA
60	2951	CTTACACCTT	GGCGGTCAAC	AATACCGGCA	ACGAACCCGT	AAGCCTCGAT
	3001	CAATTGACGG	TAGTGGAAGG	GAAAGACAAC	AAACCGCTGT	CCGAAAACCT
	3051	TAATTTTACC	CTGCAAAACG	AACACGTCGA	TGCCGGCGCG	TGGCGTTACC
	3101	AACTCATCCG	CAAAGACGCG	GAGTTCCGCC	TGCATAATCC	GGTCAAAGAA
	3151	CAAGAGCTTT	CCGACAAACT	CGGCAAGGCA	GAAGCCAAAA	AACAGGCGGA
65	3201	AAAAGACAAC	CGCAGAAAGC	TTGACGCGCT	GATTGCGGCC	GGGCGCGATG
	3251	CCGCCGAAAA	GACAGAAAGC	GTTGCCGAAC	CGGCCCGGCN	GGCAGGCGGG
	3301	GAAAATGTCT	GCATTATGCA	GGCGGAGGAA	GAGAAAAAAC	GGGTGCAGGC
	3351	GGATAAAGAC	AGCGCNTTGG	CGAAACAGCG	CGAAGCGGAA	ACCCGGCCGG
	3401	NTACCACCGC	CTTCCCCCGC	GCCCCGNGCG	CCCGCCGGGA	TTTGCCGCAA
70	3451	CCGCAGCCCC	AACCCCAACC	TCAACCCCAA	CCGCAGCGCG	ACCTGATNAG
	3501	CCGTTATGCC	AATAGCGGTT	TGAGTGAATT	TTCCGCCACG	CTCAACAGCG
	3551	TTTTCGCCGT	ACAGGACGAA	TTGACCGCG	TGTTTGCCGA	AGACCGCCGC

-365-

3601 AACGCNGTTT GGACAAGCNG CATCCGGNAC ACCAAACACT ACCGTTTCGCA  
 3651 AGATTTCGCG GCCTACCGCC AACAAACCGA CCTGCGCCAA ATCGGTATGC  
 3701 AGAAAAACCT CGGCAGCGGG CGCGTCGGCA TCCTGTTTTT GCACAACCGG  
 3751 ACCGAAAACA NCTTCGACGA CGGCATCGGC AACTCGGCAC GGCTTGCCCA  
 5 3801 CGGCGCCGTT TTCGGGCAAT ACGGCATCGG CAGGTTCGAC ATCGGCATCA  
 3851 GCACGGGCGC GGGTTTTAGC AGCGGCANTC TNTCAGACGG CATCGGAGGC  
 3901 AAAATCCGCC GCCGCGTGCT GCATTACGGC ATTCAGGCAC GATACGCGC  
 3951 CGGTTTCGGC GGATTTCGGC TCGAACCGTA CATCGGCGCA ACGCGCTATT  
 10 4001 TCGTCCAAAA AGCGGATTAC CGCTACGAAA ACGTCAATAT CGCCACCCCC  
 4051 GGTCTTGCGT TCAACCGNTA CCGNGCGGGC ATTAAGGCAG ATTATTCATT  
 4101 CAAACCGGCG CAACACATNT CCATCACNCC TTATTTNAGC CTGTCCTATA  
 4151 CCGATGCCGC TTCGGGCAAA GTCCGAACAC GCGTCAATAC CGCNGTATTG  
 4201 GCTCAGGATT TCGGCAAAAC CCGCAGTGC GAATGGGGCG TAAACGCCGA  
 4251 AATCAAAGGT TTCACGCTGT CCNTCCACGC TGCCGCCGCC AAAGGNCCGC  
 15 4301 AACTGGAAGC GCAACACAGC GCGGGCATCA AATTAGGCTA CCGCTGGTAA

This encodes a protein having amino acid sequence <SEQ ID 652>:

1 MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN  
 51 YQYYRDFAE NKGKFAVGAKD IEVYNKKGEL VGKSMKAPM IDFSVVS RNG  
 101 VAALVGDQYI VSVAHNGGYN NVDFGAEGXN PDQHRFSYQI VKRNNYKPDN  
 151 SHPYNGDXHM PRLHKFVTDA EPVEMTSDMR GNTYSDKEY PERVRIGSGH  
 201 HYWRYDDDKH GDLSYSGAWL IGGNTHMQGW GNNGVXSLSG DVRHANDYGP  
 251 MPIAGAAGDS GSPMFIYDKT NNKWLNLGVL QTGYPSGRE NGFOLIRKDW  
 301 FYDDIYRGDT HTVXFEPNRS GHFSFTSNNN GTGTVTETNE KVSNNPKLVQ  
 351 TVRLFDES LN ETDKEPVYAA GGVNQYRPRL NNGENLSFID YGNGLILSN  
 25 401 NINQGAGGLY FEGDFTVSPE NNETWQAGAV HISEDSTVTW KVNGVANDRL  
 451 SKIGKGT LHV QAKGENQSGI SVGDGT VILD QQADDK GK KQ AFSEIGLXSG  
 501 RGT VQLNADN QFNPDKLYFG FRGGRDLNGL HSLSFHRIQN TDEGAMIXXH  
 551 NATTTSTVTI TGNESITQPS GKNINRLNYS KEIAYNGWFG EKDTTKTNGR  
 601 LNLVYQPAAE DRTXLLSGGT NLGNITQTN GKLFSSGRPT PHAYNHLGSG  
 30 651 WSKMEGIPQG EIVWDNDWIX RTFKAENFHI QGGQAVISRN VAKVEGD XHL  
 701 SNHAQAVFGV APHQSH TICT RSDWTGLTNC VEXXITDDKV IASLTKTDXS  
 751 GXVXLXXXXX XXLXGXAXLX GNLSANGDTR YTVSHNATQN GNLSLVGNAQ  
 801 ATFNQATLNG NXSXSGNASF NLSNNAQNG SITLSDNAKA NVSHSALNGN  
 851 VSLADKAVFH FENSRTGQL SGSKXTALHL KDSEWTLPSG TELGNLNDN  
 35 901 ATITLNSAYR HDAAGATGX VSDTPRRRSR RSLLSVTPPT SVESRFTLT  
 951 VNGKLN XQGT FRFMSE LFGY RSDKLKLAES SEGTYTLAVN NTGNEPVSLD  
 1001 QLT VVEGKDN KPLSEN LNFT LQNEHVDAGA WRYQLIRKDG EFR LHPVKE  
 1051 QELSDKLGKA EAKQAEKDN AQSLDALIAA GRDAAEKTES VAEPARXAGG  
 1101 ENVGIMQAE EKKRVQADKD SALAKQREAE TRPXTTAFPR ARXARRDL PQ  
 40 1151 PQPQPQPQPQ PQRLXSRYA NSGLSEFSAT LNSVFAVQDE LDRVFAEDRR  
 1201 NAVWTSXIRX TKHYRSQDFR AYRQOTDLRQ IGMQKNLGSG RVGILFSHNR  
 1251 TENXFDDGIG NSARLAHGA VFGQYIGRFD IGISTGAGFS SGXLSDGIGG  
 1301 KIRRRVLHYG IQARYRAGFG GFGIEPYIGA TRYFVQKADY RYENVNIATP  
 1351 GLAFNRYRAG IKADYSFKPA QHXSITPYXS LSYTDAASGK VRTRVNTAVL  
 45 1401 AQDFGKTRSA EWGVNAEIKG FTL SXHAAA KGPQLEA QHS AGIKLGYRW\*

A transmembrane region is underlined.

ORF1-1 shows 86.3% identity over a 1462aa overlap with ORF1a:

10 20 30 40 50 60  
 orfla.pep MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDFAE N  
 50 orfl1-1 MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDFAE N  
 10 20 30 40 50 60  
 70 80 90 100 110 120  
 55 orfla.pep KGKFAVGAKDIEVYNKKGELVGKSMKAPMIDFSVVS RNGVAALVGDQYIVSVAHNGGYN  
 orfl1-1 KGKFAVGAKDIEVYNKKGELVGKSMKAPMIDFSVVS RNGVAALVGDQYIVSVAHNGGYN  
 70 80 90 100 110 120  
 60 orfla.pep NVDFGAEGXNPDQHRFSYQIVKRNNYKPDNS-HPYNGDXHMPRLHKFVTDAEPVEMTSDM  
 orfl1-1 NVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTDAEPVEMTSYM  
 130 140 150 160 170 180  
 65

-366-

		180	190	200	210	220	230
	orfla.pep	RGNTYSDKEYPERVRIGSGHHYWRDDEKKGDL--SYSGA----WLIGGNTHMQGWGNN					
	orfl-1	DGRKYIDQNNYPDRVRIGAGRQYWRSEDEPNNRESSYHIASAYSWLVGGNTFAQNGSGG					
5		190	200	210	220	230	240
	orfla.pep	GVXSLSGD-VRHANDYGPMPIAGAAGDSGSPMFIYDKTNNKWLNGVLQTGYPYSGRENG					
10	orfl-1	GTVNLGSEKIKHS-PYGFLPTGGSFGDSGSPMFIYDAQKQKWLINGVLQTGNPNPYIGKSN					
		250	260	270	280	290	
	orfla.pep	FQLIRKDWFYDDIYRGDTHTVXFEPRSNHGFSTSNNGTGTVTETNEKVSNP-KLKVQT					
15	orfl-1	FQLVRKDWFYDEIFAGDTHSVFYEPQNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRT					
		300	310	320	330	340	350
	orfla.pep	VRLFDESINETDKEPVY-AAGGVNQYRPRLNNGENLSFIDYGNGLILSNINQAGAGGLY					
20	orfl-1	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKGELILTSNINQAGAGGLY					
		360	370	380	390	400	410
	orfla.pep	FEGDFTVSPENNETWQAGGVHISEDSTVTWKVNGVANDRLSKIGKGTLLHVQAKGENQGS					
25	orfl-1	FQGDFTVSPENNETWQAGGVHISEDSTVTWKVNGVANDRLSKIGKGTLLHVQAKGENQGS					
30		420	430	440	450	460	470
	orfla.pep	SVGDGTVILDQQADDKGGKQAFSEIGLXSGRGTVQLNADNQFNPDKLYFGFRGRLDLNG					
	orfl-1	SVGDGTVILDQQADDKGGKQAFSEIGLVSGRGTVQLNADNQFNPDKLYFGFRGRLDLNG					
35		480	490	500	510	520	530
	orfla.pep	HSLSFHRIQNTDEGAMIXXHNATTTSTVTITGNESITQPSGKNINRLNYSKEIAYNGWFG					
40	orfl-1	HSLSFHRIQNTDEGAMIVNHNQDKESTVTITGNKDIAT-TGNN-NSLDSKKEIAYNGWFG					
		540	550	560	570	580	590
	orfla.pep	EKDTTKTNGRLNLVYQPAEDRTXLLSGGTNLNGNITQTNGKLEFFSGRPTPHAYNHLGSG					
45	orfl-1	EKDTTKTNGRLNLVYQPAEDRTLLSGGTNLNGNITQTNGKLEFFSGRPTPHAYNHLNDH					
		600	610	620	630	640	650
	orfla.pep	WSKMEGIPQGEIVWDNDWIXRTFKAENFHIQGGQAVISRNVAKVEGDHLSNHAQAVFGV					
50	orfl-1	WSQKEGIPRGEIVWDNDWINRTFKAENFQIKGGQAVVSRNVAKVKGDWHLNHAQAVFGV					
		660	670	680	690	700	710
	orfla.pep	APHQSHITICTRSDWTGLTNCVEXXITDDKVIASLTKTDXSGXVXLXXXXXXLXGAXLX					
55	orfl-1	APHQSHITICTRSDWTGLTNCVEKITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLN					
60		720	730	740	750	760	770
	orfla.pep	GNLSANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNXSXSGNASFNLSNNAQNG					
	orfl-1	GNLSANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSHNAVQNG					
65		780	790	800	810	820	830
	orfla.pep	SLTSLDNAKANVSHSALNGNVSLADKAVFHFENSRTFGQLSGSKXTALHLKDSEWTLPSG					
70	orfl-1	SLTSLDNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSG					
		840	850	860	870	880	890

-367-

		900	910	920	930	940
	orfla.pep	TELGNLNL	DNATITL	NSAYRHD	AAGAQTG	XVSDTPR
5	orfl-1	TELGNLNL	DNATITL	NSAYRHD	AAGAQTG	SATDAPR
		900	910	920	930	940
		950	960	970	980	990
	orfla.pep	TLTVNGK	LNQGTFR	FMSELF	GYRSDK	LKLAES
10	orfl-1	TLTVNGK	LNQGTFR	FMSELF	GYRSDK	LKLAES
		960	970	980	990	1000
		1010	1020	1030	1040	1050
	orfla.pep	KDNKPL	SENLF	NFTLQ	NEHVD	DAGAWRY
15	orfl-1	KDNKPL	SENLF	NFTLQ	NEHVD	DAGAWRY
		1020	1030	1040	1050	1060
		1070	1080	1090	1100	1110
	orfla.pep	KDNAQSL	DALIA	AGRDAA	EKTESV	AEPARX
20	orfl-1	KDNAQSL	DALIA	AGRDAA	EKTESV	AEPARX
		1080	1090	1100	1110	1120
		1130	1140	1150	1160	1170
	orfla.pep	EAETRP	XTTAF	PRARX	ARRDL	PQPQP
	orfl-1	EAETRP	ATTAF	PRARR	ARRDL	PQLQP
		1140	1150	1160	1170	1180
		1190	1200	1210	1220	1230
	orfla.pep	QDELDR	VFAED	RRNAV	WTSXIR	XTKHYS
35	orfl-1	QDELDR	VFAED	RRNAV	WTSXIR	XTKHYS
		1200	1210	1220	1230	1240
		1250	1260	1270	1280	1290
	orfla.pep	HNRTEN	XFDG	IGNSA	RLAHG	AVFGQ
40	orfl-1	HNRTEN	XFDG	IGNSA	RLAHG	AVFGQ
		1260	1270	1280	1290	1300
		1310	1320	1330	1340	1350
	orfla.pep	HYGIQ	ARYR	AGFG	GFGIE	PYIGAT
45	orfl-1	HYGIQ	ARYR	AGFG	GFGIE	PYIGAT
		1320	1330	1340	1350	1360
		1370	1380	1390	1400	1410
	orfla.pep	KPAQH	XSI	TPYX	SLSYT	DAASG
50	orfl-1	KPAQH	XSI	TPYX	SLSYT	DAASG
		1380	1390	1400	1410	1420
		1430	1440	1450		
	orfla.pep	AAAKG	PQLE	AQHS	AGIKL	GYRWX
	orfl-1	AAAKG	PQLE	AQHS	AGIKL	GYRWX
		1440	1450			

Homology with adhesion and penetration protein hap precursor of *H.influenzae* (accession number P45387)

Amino acids 23-423 of ORF1 show 59% aa identity with hap protein in 450aa overlap:

	orfl	23	FXAAYLAICLSFGILPQAWAGHTYFGINYQYYRDF	82
			F +L C+S GI QAWAGHTYFGI+YQYYRDF	
65	hap	6	FRLNFLTACVSLGIASQAWAGHTYFGIDYQYYRDF	65
			ENKKGKFAVGAKDIEVYNKKGELVG	
	orfl	83	KSMTKAPMIDFSVVSRRNGVAALVGVQYIVSVAHNGGYNNVDFGAEGXNIXDQXRTYKIV	142
			SMTKAPMIDFSVVSRRNGVAALVG QYIVSVAHNGGYN+VDFGAEG N DQ R TY+IV	
	hap	66	TSMTKAPMIDFSVVSRRNGVAALVGDQYIVSVAHNGGYNDVDFGAEGRN-PDQHRFTYQIV	124



-368-

5  
10  
15  
20  
25

orf1 143 KRNNYKAGTKGHPYGGDYHMPRLHKKXVTD AEPVEMTSYMDGRKYIDQNNYPDRVRIGAGR 202  
KRNNY+A + HPY GDYHMPRLHK VT+AEPV MT+ MDG+ Y D+ NYP+RVRIG+GR  
hap 125 KRNNYQAWERKHPYDGDYHMPRLHKKFVTEAEPVGMTTNMDGKVYADRENYPERVERIGSGR 184

orf1 203 QYWRSEDEPNRESSYHIA----- 222  
QYWR+D+DE N SSY+++  
hap 185 QYWRTDKDEETNVHSSYYVSGAYRYLTAGNTHQTSGNGNGTVNLSGNVVSPNHYGPLPTG 244

orf1 223 -----SGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGFQLVKRDWYFDEIFAGDTHSVF 277  
SGSPMFIYDA+K++WLIN VLQTG+P+ G+ NGFQL+R++WFY+E+ A DT SVF  
hap 245 GSKGDSGSPMFIYDAKKKQWLINAVLQTGHPFFGRGNGFQLIREEFYFNEVLAVDTPSVF 304

orf1 278 --YEPRQNGKYSFNDDNNGTGKIN-AKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHA 334  
Y P NG YSF +N+GTGK+ + + + + TV+LFN SL++TA+E V A  
hap 305 QRYIPPINGHYSFVSNNDGTGKLTLTRPSKDGSKAKSEVGTVKLFNPSLNQTAKEHV-KA 363

orf1 335 AGGVNSYRPRLNNGENISFIDEKGKELILTSNINQGAGGLYFQGDFTV-SPENNETWQGA 393  
A G N Y+PR+ G+NI D+GKG L + +NINQGAGGLYF+G+F V +NN TWQGA  
hap 364 AAGYNIYQPRMEYKKNYILGDQKGKTLTIENNINQGAGGLYFEGNFVVKGKQNNITWQGA 423

orf1 394 GVHISEDSTVTWKVNGVANDRLSKIGKGT 423  
GV I +D+TV WKV+ NDRLSKIG GTL  
hap 424 GVSIGQDATVEWKVHPENDRLSKIGIGTL 453

Amino acids 715-1011 of ORF1 show 50% aa identity with hap protein in 258aa overlap:

30  
35  
40  
45

Orf1 41 DTRYTVSHNATQ-NGNXSLVXNAQATFNQ-ATLNGNTSASGNASFNLSDHAVQNGSLTSL 98  
DT+ S TQ NG+ +L NA + A LNGN + ++ F LS++A Q G++ LS  
hap 733 DTKVINSIPITQINGSINLTNNATVNIHGLAKLNGNVTIDHSQFTLSNNATQTGNIKLS 792

orf1 99 GNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGXELGN 158  
+A A V+++ LNGNV L D A F ++S F QI G KDT + L+++ WT+PS L N  
hap 793 NHANATVNNATLNGNVHLTDSAQFSLKNSHFWHQIQGDKDTTVTLENATWTMPSDTTLQN 852

orf1 159 LNLDNATITLNSAYRHDAAGAQTGSATDAPXXXXXXXXXXLLXVTPPTSVESEFNTLTVN 218  
L L+N+T+TLNSAY + S+ +AP L T PTS E RFNTLTVN  
hap 853 LTLNNSTVTLNSAY-----SASSNNAPRHRS-----LETETTP TSAEHRFNTLTVN 899

orf1 219 GKLNQOGTFRFMSELFYGRSDKLKLAESSEGTYYTLAVNNTGNEPASLEQLTVVEGKDNKP 278  
GKL+GQGTFF+ S LFGY+SDKLKL+ +EG YTL+V NTG EP +LEQLT++E DNKP  
hap 900 GKLSGQGTFOFTSSLFYKSDKLKLSNDAEGDYTLVRNTGKEPVTLEQLTLIESLDNKP 959

orf1 279 LSENLFNFTLQNEHVDAGA 296  
LS+ L FTL+N+HVDAGA  
hap 960 LSDKLKFTLENDHVDAGA 977

Amino acids 1192-1450 of ORF1 show 41% aa identity with hap protein in 259aa overlap:

50  
55  
60  
65

Orf1 1 LDRVFAEDRRNAVWTS GIRDTHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNR 60  
LDR+F + ++AVWT+ +D + Y S FRAY+Q+T+LRQIG+QK L +GR+G +FSH+R  
hap 1135 LDRLFVDQAQSAVWTNIAQDKRRYDSDAFRAYQOKTNLRQIGVQKALANGRIGAVFSHSR 1194

orf1 61 TENTFDDGIGNSARLAHGA VFGQY GIDRFYXXXXXXXXXXXXXXXXXIGKXRRRVLHYG 120  
++NTFD+ + N A L + F QY K R+ ++YG  
hap 1195 SDNTFDEQVKNHATLTMMSGFAQYQWGD LQFGVNVGTGISASKMAEEQSRKIHRAKAINYG 1254

orf1 121 IQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPA 180  
+ A Y+ G GI+P+ G RYF+++ +Y+ E V + TP LAFNRY AGI+ DY+F P  
hap 1255 VNASYQFRLGQLGIQPYFGVNRYFIERENYQSEEVVRKTPSLAFNRYNAGIRVDYTFPT 1314

orf1 181 QHISITPYLSLSYTD AASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAA 240  
+IS+ PY ++Y D ++ V+T VN VL Q FG+ E G+ AEI F +S + +  
hap 1315 DNISVKPYFFVNYVDVSNANVQTTVNLTVLQQPFGRYWQKEVGLKAEILHFQISAFISKS 1374

orf1 241 KGPQLEAQHSAGIKLGYRW 259  
+G QL Q + G+KLG YRW  
hap 1375 QGSQLGKQONVGVKLG YRW 1393

Homology with a predicted ORF from *N.gonorrhoeae*

The blocks of ORF1 show 83.5%, 88.3%, and 97.7% identities in 467, 298, and 259 aa overlap, respectively with a predicted ORF (ORF1ng) from *N.gonorrhoeae*:

5	orf1.pep	MKTTDKRTTETHRKAPKTGRIRFXAAYLAICLSFGILPQAWAGHTYFGINYQYYRDAEN	60
	orf1ng	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDAEN	60
10	orf1.pep	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALVGQYIVSVAHNGGYN	120
	orf1ng	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALAGDQYIVSVAHNGGYN	120
15	orf1.pep	NVDFGAEGXNIXDQXRXYKIVKRNNYKAGTKGHPYGGDYHMPRLHKXVTDAPVEMTSY	180
	orf1ng	NVDFGAEGSN-PDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTDAPVEMTSY	179
20	orf1.pep	MDGRKYIDQNNYPDRVRIGAGRQYWRSDDEPNRESSYHIAS-----	223
	orf1ng	MDGWKYADLNKYPDRVRIGAGRQYWRSDDEPNRESSYHIASAYSWLVGGNTFAQNGSG	239
25	orf1.pep	-----GSPMFIYDAQKQKWLINGVLOTGNPYIGKSNG	255
	orf1ng	GGTVNLGSEKIKHSPYGFLLPTGGSGSGSPMFIYDAQKQKWLINGVLOTGNPYIGKSNG	289
30	orf1.pep	<u>FQLVRKDWFYDEIFAGDTHSVFYEP</u> RQNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRT	315
	orf1ng	<u>FQLVRKDWFYDEIFAGDTHSVFYEP</u> HQNGKYFFNDNNNGAGKIDAKHKHYSLPYRLKTRT	359
35	orf1.pep	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKGELILTSNINQAGAGGLY	375
	orf1ng	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDKGGKELILTSNINQAGAGGLY	
40	orf1.pep	FQGDFTVSPENNETWQAGGVHISEDSTVTWKVNGVANDRLSKIGKGT	422
	orf1ng	FEGNFTVSPKNNETWQAGGVHISDGSTVTWKVNGVANDRLSKIGKGTLLVQAKGENQGSV	479
45	orf1.pep	//	
	orf1ng	DKVTASLTKTDISGNVDLADHAHLNLTGLA	744
50	orf1.pep	FGVAPHQSHITICRSWDTGLTSCTEKTITDDKVIASLSKTDVRGNVSLADHAHLNLTGLA	774
	orf1ng	TLNGNLSANGDTR-YTVSHNATQNGNXSLVXNAQATFNQATLNGNTSASGNASFNLSNNA	803
55	orf1.pep	TFNGNL-VQAETRTIRLRANATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLSNNA	833
	orf1ng	VQNGSLTSLGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWT	863
60	orf1.pep	VQNGSLTSLDNKANVSHSALNGNVSLADKAVFHFENSRTGKISGGKDTALHLKDSEWT	893
	orf1ng	LPSGXLGNLNLNLDNATITLNSAYRHAAGAQTGSATDAPRRRSRRSRLXVTPPTSVE	923
65	orf1.pep	LPSGTELGNLNLNLDNATITLNSAYRHAAGAQTGSAADAPRRRSRRS---LLSVTPPTSVE	950
	orf1ng	SRFNTLTVNGKLNQGTFRFMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLT	983
70	orf1.pep	SRFNTLTVNGKLNQGTFRFMSELFGYRSGKLKLAESSEGTYTLAVNNTGNEPVSLEQLT	1010
	orf1ng	VVEGKDNKPLSENLFNLQNEHVDAGAW	1011
75	orf1.pep	VVEGKDNTPLSENLFNLQNEHVDAGAWRYQLIRKDGFEFRLHNPVKEQELSDKLKAGET	1070
	orf1ng	//	
80	orf1.pep	LDRVFAEDRRNAVWTSIGIRDTKHYRSQDFR	1211
	orf1ng	PQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEDRRNAVWTSIGIRDTKHYRSQDFR	1239
85	orf1.pep	AYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENTFDDGIGNSARLAHGAVFQYIGIDRFY	1271
	orf1ng	AYRQQTDLRQIGMQKNLGSGRVGILFSHNRTGNTFDDGIGNSARLAHGAVFQYIGIRFD	1299

-370-

```

orfl.pep      IGISAGAGFSSGSLSDGIGXKRRRLVHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY 1331
orflng       IGISAGAGFSSGSLSDGIRGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY 1359

5  orfl.pep      RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDASGKVRTRVNTAVL 1391
orflng       RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDASGKVRTRVNTAVL 1419

10 orfl.pep      AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGPQLEAQSAGIKLGYRW 1440
orflng       AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGPQLEAQSAGIKLGYRW 1468

```

The complete length ORF1ng nucleotide sequence was identified <SEQ ID 653>:

```

1  ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCTAA
51 AACC GGCCGC ATCCGCTTCT CGCCCGCTTA CTTAGCCATA TGCCTGTCGT
15 101 TCGGCATTCT GCCCAAGCC CGGGCGGGAC ACACTTATTT CGGCATCAAC
151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
201 GCGGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAAT
251 CGATGACGAA AGCCCCGATG ATTGATTTTT CTGTGGTATC GCGTAACGGC
301 GTGGCGGCAT TGGCGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG
20 351 CGGCTATAAC AATGTTGATT TTGGTGC GGAAGCAAT CCCGATCAGC
401 ACCGCTTTTC TTACCAAATT GTGAAAAGAA ATAATTATAA AGCAGGGACT
451 AACGGCCATC CTTATGGCGG CGATTATCAT ATGCCGCGTT TGCACAAATT
501 TGTCACAGAT GCAGAACCTG TTGAGATGAC CAGTTATATG GATGGGTGGA
551 AATACGCTGA TTTAAATAAA TACCCTGATC GTGTTCTGAAT CGGAGCAGGC
25 601 AGACAATATT GGCGGTCTGA TGAAGACGAA CCCAATAACC GCGAAAGTTC
651 ATATCATATT GCAAGCGCAT ATTCTTGGCT CGTCGGTGGC AATACCTTTG
701 CACAAAATGG ATCAGGTGGT GGCACAGTCA ACTTAGGTAG CGAAAAAATT
751 AAACATAGCC CATATGGTTT TTTACCAACA GGAGGCTCAT TTGGCGACAG
801 TGGCTCACA ATGTTTATCT ATGATGCCCA AAAGCAAAAG TGGTTAATTA
30 851 ATGGGGTATT GCAAACAGGC AACCCTATA TAGGAAAAG CAATGGCTTC
901 CAGCTAGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC
951 CCATTCAGTA TTCTACGAAC CACATCAAAA TGGGAAATAC TTTTTTAACG
1001 ACAATAATA TGGCGCAGGA AAAATCGATG CCAAACATAA AACTATTCT
1051 CTACCTTATA GATTAATAAC ACGAACCCTT CAATTGTTTA ATGTTTCTTT
35 1101 ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGGGTCAACA
1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCTTT TATTGACAAA
1201 GGAAAAGGTG AATTGATACT TACCAGCAAC ATCAACCAAG GCGCGGGCGG
1251 TTTGTATTTT GAGGGTAATT TTACGGTCTC GCCTAAAAAC AACGAAACGT
1301 GGCAAGGCGC GGGCGTTCAT ATCAGTGATG GCAGTACCGT TACTTGGAAA
40 1351 GTAAACGGCG TGGCAAACGA CCGCCTGTCC AAAATCGGCA AAGGCACGCT
1401 GCTGGTTCAA GCCAAAGGGG AAAACCAAGG CTCGGTCAGC GTGGGCGACG
1451 GTAAAGTCAT CTTAGATCAG CAGGCGGACG ATCAAGGCAA AAAACAAGCC
1501 TTTAGTGAAA TCGGCTTGGT CAGCGGCAGG GGGACGGTGC AACTGAATGC
1551 CGATAATCAG TTCAACCCCG ACAAACTCTA TTTCCGGCTT CCGCGCGGAC
45 1601 GTTTGGATTT GAACGGGCAT TCGCTTTCGT TCCACCGCAT TCAAAATACC
1651 GATGAAGGGG CGATGATTGT CAACCACAAT CAAGACAAAG AATCCACCGT
1701 TACCATTACA GGCAATAAAG ATATTACTAC AACCGGCAAT AACAACAAT
1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAGAT
1801 GCAACCAAAA CGAAGCGGCG GCTCAATCTG AATTACCAAC CGGAAGAAGC
50 1851 GGATCGCACT TTACTGCTTT CCGGCGGAAC AAATTTAAAC GGCAATATCA
1901 CGCAACAAA CGGCAAACTG TTTTTCAGCG GCAGACCGAC ACCGCACGCC
1951 TACAATCATT TAGGAAGCGG GTGGTCAAAA ATGGAAGGTA TCCCACAAGG
2001 AGAATTCGTG TGGGACAACG ATTGGATCGA CCGCACATTT AAAGCGGAAA
2051 ACTTCCATAT TCAGGGCGGA CAAGCGGTGG TTTCCCGCAA TGTTCGCAAA
55 2101 GTGGAAGGCG ATTGGCATTT AAGCAATCAC GCCAAGCAG TTTTCGGTGT
2151 CGCACCAGAT CAAAGCCACA CAATCTGTAC ACGTTCGGAC TGGACGGGTC
2201 TGACAAGTTG TACCGAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA
2251 TTGAGCAAGA CCGACATCAG AGGCAATGTC AGCCTTGCCG ATCAGCTCA
2301 TTTAAATCTC ACAGGACTTG CCACACTCAA CGGCAATCTT AGTGCAGCG
60 2351 GAGACACGCA CTATACGGTT ACGCGCAACG CCACCCAAA CGGCAACCTC
2401 AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAACCGG
2451 CAACACATCG GCTTCGGACA ATGCTTCATT TAATCTAAGC AACACGCGG
2501 TACAAAACGG CAGTCTGACG CTTTCCGACA ACGCTAAGGC AAACGTAAGC
2551 CATTCCGCAC TCAACGGCAA TGTCTCCCTA GCCGATAAGG CAGTATTCCA
65 2601 TTTTGA AAC AGCCGCTTTA CCGGAAAAAT CAGCGGCGGC AAGGATACGG
2651 CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTGCGG CACGGAATTA
2701 GGCAATTTAA ACCTTGACAA CGCCACCATT AACTCAATT CCGCCTATCG
2751 ACACGATGCG GCAGGCGCGC AAACCGGCAG TCGGCGCAGT GCGCGCGCC
2801 GCCGTTCCGC CCGTTCCCTA TTATCCGTTA CGCCGCCAAC TTCGGCAGAA
70 2851 TCCCGTTTCA ACACGCTGAC GGTAACGGC AAATTGAACG GTCAGGGAAC

```

-371-

2901 ATTCCGCTTT ATGTCGGAAC TCTTCGGCTA CCGCAGCGGC AAATTGAAGC  
 2951 TGGCGGAAAG TTCCGAAGGC ACTTACACCT TGGCTGTCAA CAATACCGGC  
 3001 AACGAACCCG TAAGTCTCGA GCAATTGACG GTAGTGGAAG GAAAAGACAA  
 3051 CACACCGCTG TCCGAAATC TTAATTTTAC CCTGCaaaAc gaacacgtcg  
 5 3101 atgccggcgc atggCGTTAT CAGCTTATCC gcaaagacgG CGAGTTCCGc  
 3151 CTGCATAATC CCGTCAAAGA ACAAGAGCTT TCCGACAAAC TCGGCAAGgc  
 3201 gggagaaACA GAggcccctT TGACGGCAAA ACAGGCacaA CTTGCCGCCA  
 3251 AAcacaggc ggaAAAAGAC AACgcgcaaa gccttgAcgc gctgattgcg  
 10 3301 gCggggcgca atgccaccga AAAGGCagaa agtggtgccc aaccgCCCCG  
 3351 GCAGGCAGGC GGGGAAAAtg ccgGCATTAT GCAGGCGGAG GAAGAGAAAA  
 3401 AACGGGTGCA GGCGGATAAA GACACCGCCT TGGCGAAACA GCGCGAAGCG  
 3451 GAAACCCGGC CGGCTACCAC CGCCTTCCCC CGCGCCCGCC GCGCCCGCCG  
 3501 GGATTTGCCG CAACCGCAGC CCCAACCGCA ACCCCAACCG CAGCGCGACC  
 15 3551 TGATCAGCCG TTATGCCAAT AGCGGTTTGA GTGAATTTTC CGCCACGCTC  
 3601 AACAGCGTTT TCGCCGTACA GGACGAATTG GACCGCGTGT TTGCCGAAGA  
 3651 CCGCCGCAAC GCCGTTTGGG CAAGCGGCAT CCGGGACACC AAACACTACC  
 3701 GTTCGCAAGA TTTCCGCGCC TACCGCCAAC AAACCGACCT GCGCCAATC  
 3751 GGTATGCAGA AAAACCTCGG CAGCGGGCGC GTCGGCATCC TGTTTTCGCA  
 20 3801 CAACCGGACC GGAACACCT TCGACGACGG CATCGGCAAC TCGGCACGGC  
 3851 TTGCCACGG TGCCGTTTTT GGGCAATACG GCATCGGCAG GTTCGACATC  
 3901 GGCATCAGCG CGGGCGCGGG TTTTAGTAGC GGCAGCCTTT CAGACGGCAT  
 3951 CAGAGGCAAA ATCCGCGGCC GCGTGCTGCA TTACGGCATT CAGGCAAGAT  
 4001 ACCGCGCAGG TTTCGGCGGA TTCGGCATCG AACCGCACAT CGGCGCAACG  
 25 4051 CGTATTTTCG TCCAAAAGC GGATTACCGA TACGAAAACG TCAATATCGC  
 4101 CACCCCGGGC CTTCGATTCA ACCGCTACCG CGCGGGCATT AAGGCAGATT  
 4151 ATTCATTCAA ACCGGCGCAA CACATTTCCA TCACGCCTTA TTTGAGCCTG  
 4201 TCTATACCG ATGCCGCTTC CGGCAAAGTC CGAACGCGC TCAATACCGC  
 4251 CGTATTGGCG CAGGATTTTC GAAAACCCG CAGTGCGGAA TGGGGCGTAA  
 30 4301 ACGCCGAAAT CAAAGTTTC ACGCTGTCCC TCCACGCTGC CGCCGCAAG  
 4351 GGGCCGCAAT TGGAAGCGCA GCACAGCGCG GGCATCAAAT TAGGCTACCG  
 4401 CTGGTAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 654>:

1 MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPOA RAGHTYFGIN  
 35 51 YQYYRDFAEK KGFVAVGAKD IEVYNKKGEL VGKSMKAPM IDFSVVRNG  
 101 VAALAGDQYI VSAHNGGYN NVDFGAEGSN PDQHRFSYQI VKRNNYKAGT  
 151 NGHYPGGDYH MPRLHKFVTD AEPVEMTSYM DGWKYADLNK YPDRVRIGAG  
 201 RQYWRSEDEH PNNRESSYHI ASAYSWLVG NTFAQNGSGG GTVNLGSEKI  
 251 KHSYPYGLPT GSGFGDSGSP MFIYDAQKQK WLINGVLOTG NPYIGKSNFG  
 301 QLVRKDWFYD EIFAGDTHSV FYEPHQNGKY FFNDNNNGAG KIDAKHKHYS  
 40 351 LPYRLKTRTV QLFNVSLSET AREPVYHAAG GVNSYRPRLN NGENISFIDK  
 401 GKGLILTSN INQGAGGLYF EGNFTVSPKN NETWQAGAVH ISDGSVTWTK  
 451 VNGVANDRLS KIGKGTLLVQ AKGENQGSVS VGDGKVILDQ QADDQGGKQA  
 501 FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSFHRIQNT  
 551 DEGAMIVNHN QDKESTVTIT GNKDITTTGN NNNLDSKKEI AYNGWFGEKD  
 45 601 ATKTNGGLNL NYPPEADRT LLLSGGTNLN GNITQTNGKL FFSGRPTPHA  
 651 YNHLGSGWSK MEGIPQGEIV WDNDWIDRTF KAENFHIQGG QAVVSRNVAK  
 701 VEGDWHLNSH AQAVFGVAPH QSHTICTRSD WTGLTSCTEK TITDDKVIAS  
 751 LSKTDVRGNV SLADHAHLNL TGLATFNGNL VQAEFRTIRL RANATQNGNL  
 801 SLVGNAQATF NQATLNGNTS ASDNASFNLS NNAVQNGSLT LSDNAKANVS  
 50 851 HSALENGVSL ADKAVFHFEN SRFTGKISGG KDTALHLKDS EWTLPSTGTEL  
 901 GNLNLDNATI TLNSAYRHDA AGAQTGSAAD APRRRSRRSL LSVTPPTSSE  
 951 SRENTLTVNG KLNGQGTFRF MSELFYGRSG KLKLAESSEG TYTLAVNNTG  
 1001 NEPVSLLEQLT VVEGKDNTPL SENLNFTLQN EHVDAGAWRY QLIRKDGFR  
 55 1051 LHPVKEQEL SDKLGKAGET EAALTAKQAQ LAAKQQAQEKD NAQSLDALIA  
 1101 AGRNATEKAE SVAEPARQAG GENAGIMQAE EEKRVQADK DTALAKQREA  
 1151 ETRPATTAFF RARRARRDLP QPQPQPQPQ QRDLSRYAN SGLSEFSATL  
 1201 NSVFAVQDEL DRVFAEDRRN AVWTSGIRDY KHYRSQDFRA YRQQTDLRQI  
 1251 GMQKNLGSGR VGILFSHNRT GNTFDDGIGN SARLAHGAFF GQYIGRFDI  
 60 1301 GISAGAGFSS GSLSDGIRGK IRRRVLYHYG QARYRAGFGG FGIEPHIGAT  
 1351 RYFVQKADYR YENVNIATPG LAFNRYRAGI KADYSFKPAQ HISITPYLSL  
 1401 SYTDAASGKV RTRVNTAVLA QDFGKTRSAE WGVNAEIKGF TSLSLHAAAAK  
 1451 GPQLEAQHSA GIKLGYRW\*

Underlined and double-underlined sequences represent the active site of a serine protease (trypsin family) and an ATP/GTP-binding site motif A (P-loop).

65 ORF1-1 and ORF1ng show 93.7% identity in 1471 aa overlap:

-372-

		10	20	30	40	50	60
	orf1-1.pep	MKT	TDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYYQYR	DF	AE	N	
	orf1-1						
5	orf1-1	MKT	TDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYYQYR	DF	AE	N	
		10	20	30	40	50	60
	orf1-1.pep	70	80	90	100	110	120
	orf1-1	KGKFAVGA	KDIEVYNKKGELVGKSM	TKAPMIDFSVVS	RNGVAALVGDQYIV	SVAHN	GGYN
10	orf1-1	KGKFAVGA	KDIEVYNKKGELVGKSM	TKAPMIDFSVVS	RNGVAALAGDQYIV	SVAHN	GGYN
		70	80	90	100	110	120
	orf1-1.pep	130	140	150	160	170	180
	orf1-1	NVDFGAEG	RNPdqhrfTYKIVKR	NNYKAGTKGHPYGG	DYHMPRLHKFVTDAE	PVEMTSY	
15	orf1-1	NVDFGAEG	RNPdqhrfSYQIVKR	NNYKAGTNGHPYGG	DYHMPRLHKFVTDAE	PVEMTSY	
		130	140	150	160	170	180
	orf1-1.pep	190	200	210	220	230	240
	orf1-1	DGRKYIDQ	NNYPDRVRIGAGRQY	WRSDEDEPN	NRESSYHIASAYS	SWLVGGNTFAQ	NGSGG
20	orf1-1	DGRKYIDQ	NNYPDRVRIGAGRQY	WRSDEDEPN	NRESSYHIASAYS	SWLVGGNTFAQ	NGSGG
		190	200	210	220	230	240
	orf1-1.pep	250	260	270	280	290	300
	orf1-1	GTVNLGSE	KIKHSPYGF	LPTGGSGFGDSG	SPMFIYDAQKQK	WLINGVLQ	TGNPYIGKSNGF
25	orf1-1	GTVNLGSE	KIKHSPYGF	LPTGGSGFGDSG	SPMFIYDAQKQK	WLINGVLQ	TGNPYIGKSNGF
		250	260	270	280	290	300
30	orf1-1.pep	310	320	330	340	350	360
	orf1-1	QLVRKDW	FYDEIFAGDTHSV	FYEPQNGKYSF	NDDNNGTGKINAK	HEHNSLPN	RLKTRTV
35	orf1-1	QLVRKDW	FYDEIFAGDTHSV	FYEPQNGKYSF	NDDNNGAGKIDAK	HKHYSLPY	RLKTRTV
		310	320	330	340	350	360
	orf1-1.pep	370	380	390	400	410	420
	orf1-1	QLFNVSLS	ETAREPVYHAAGG	VNSYRPLNNGE	NISFIDEGKGEL	ILTSNINQ	GAGGLYF
40	orf1-1	QLFNVSLS	ETAREPVYHAAGG	VNSYRPLNNGE	NISFIDKGKEL	ILTSNINQ	GAGGLYF
		370	380	390	400	410	420
	orf1-1.pep	430	440	450	460	470	480
	orf1-1	QGDFTVSP	ENNETWQAGVHI	SEDSTVTWKV	NGVANDRLSKIG	KGTLLHVQAK	GENQGSIS
45	orf1-1	EGNFTVSP	PKNNETWQAGVHI	SDGSTVTWKV	NGVANDRLSKIG	KGTLLVQAK	GENQGSVS
		430	440	450	460	470	480
	orf1-1.pep	490	500	510	520	530	540
	orf1-1	VGDGTVIL	DQQADDQKQKQAF	SEIGLVSGRG	TVQLNADNQFNP	DKLYFGFRG	GRDLN
50	orf1-1	VGDGTVIL	DQQADDQKQKQAF	SEIGLVSGRG	TVQLNADNQFNP	DKLYFGFRG	GRDLN
		490	500	510	520	530	540
	orf1-1.pep	550	560	570	580	590	600
	orf1-1	SLSFHRIQ	NTDEGAMIVNHNQ	DKESTVTITGN	KDIATTGNNNSL	DSKKEIAYNG	WFGEKD
55	orf1-1	SLSFHRIQ	NTDEGAMIVNHNQ	DKESTVTITGN	KDITTTGNNNSL	DSKKEIAYNG	WFGEKD
		550	560	570	580	590	600
60	orf1-1.pep	610	620	630	640	650	660
	orf1-1	TTKTNGRL	NLVYQPAEDRTLL	LSGGTNLNGNIT	QTNGKLFFSGR	PTPHAYNHL	NDHWSQ
65	orf1-1	ATKTNGRL	NLVYQPEEDRTLL	LSGGTNLNGNIT	QTNGKLFFSGR	PTPHAYNHL	SGWSK
		610	620	630	640	650	660
	orf1-1.pep	670	680	690	700	710	720
	orf1-1	KEGIPRGE	IVDNDWINRTFKA	ENFQIKGGQAV	VS	RNVAKVKG	DWHLNHAQAVFGVAPH
70	orf1-1	MEGIPQGE	IVDNDWIDRTFKA	ENFHIQGGQAV	VS	RNVAKVEGDWHLNHAQAVFGVAPH	
		670	680	690	700	710	720

-373-

		730	740	750	760	770	780
	orf1-1.pep	QSHTICTRSDWTGLTNCVEKTIITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLNGNL					
	orf1ng-1	: : :       :     : : :					
5		730	740	750	760	770	780
	orf1-1.pep	790	800	810	820	830	840
	orf1ng-1	SANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLT					
10		790	800	810	820	830	840
	orf1-1.pep	: :					

In addition, ORF1ng shows 55.7% identity with hap protein (P45387) over a 1455aa overlap:

10

15

20

25

30

35

40

50

55

60

65

70

10

15

20

25

30

35

40

45

50

55

65

70



-376-

		1320	1330	1340	1350	1360	1370
	orf1ng-1.pep	SLSDGIRGKIRRRVLHYGIQARYAGFGGFGIEPHIGATRYFVQKADYRYENVNIA TPGL					
	p45387	:::   : ::: :::    :   :  : ::: ::: ::: ::: ::: ::: ::: :::					
5		1240	1250	1260	1270	1280	1290
		KMAEEQSRKIHRKAINYGVNASYQFRLGQLGIQPYFGVNRYFIERENYQSEEV RVKTPSL					
		1380	1390	1400	1410	1420	1430
	orf1ng-1.pep	AFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAA SGKVRTRVNTAVLAQDFGKTRSAEW					
	p45387	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
10		1300	1310	1320	1330	1340	1350
		AFNRYNAGIRVDYTF TPTDNISVKPYFFVNYVDVSNANVQT TVNLTVLQQPFGRYWQKEV					
		1440	1450	1460	1469		
	orf1ng-1.pep	GVNAEIKGFTLSLHAAAAGPQLEA QHSAGIKLGYRWX					
	p45387	::      :   : ::      ::: ::: ::: ::: ::: ::: ::: ::: :::					
15		1360	1370	1380	1390		
		GLKAEILHFQISAFISKSQGSQ LGKQQNVGVKLG YRW					

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 20 Example 78

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 655>:

25  
30

```

      1  ..AAGGTGTGGC AATTTGTCGA AGA.CCGCTG CGTGCCGTCG TGCCTGCCGA
     51  CAGTTTTTGAA CCGACCGCGC AAAAATTGAA CCTGTTTAAG GCGGGTGCGG
    101  CAACCAATTTT GTTTTATGAA GATCAAAATG TCGTCAAAGG TTTGCAGGAG
    151  CAGTTCCTCGT CTTATGCCGC TAACTTCCCG GTTTGGCGGg ATCAGGCAAA
    201  CGCGATGGTG  CAGTATGCCG TTTGGACGAC ACTTGGCCGC GTCCGGCGTAG
    251  GTGCAAACCT  GCAACATTAC AATCCCTTGC CCGATGCGGC GATTGCCAAA
    301  GCGTGGAAATA TCCCCGAAAG CTGTTTGTTG CGCGACAAA TGGTTATCGG
    351  CGGTATTGAA  GGGCGGCGAC GTGAAAAGAC CTTTGAACCC GTTGCAGAAC
    401  GTTTGAAAGT  GTTCGCGCGA TAA

```

This corresponds to the amino acid sequence <SEQ ID 656; ORF6>:

```

1  ..KVWQFVEXPL RAVVPADSFE PTAQKLNLFK AGAATILFYE DQNVVKGLQE
51  QFPAYAANFP VWADQANAMV QYAVWTTLAA VGVGANLQHY NPLPDAAIAK
101 AWNIPENWLL RAQMVIIGGIE GAAGEKTFEP VAERLKVFGA *

```

35 Further sequence analysis revealed a further partial DNA sequence <SEQ ID 657>:

40

1	..CTGCGTGCCG	TCGTGCCTGC	CGACAGTTTT	GAACCGACCG	CGCAAAAATT
51	GAACCTGTTT	AAGGCGGGTG	CGCAACCAT	TTTGTTTTAT	GAAGATCAAA
101	ATGTCGTCAA	AGGTTTGCAG	GAGCAGTTCC	CTGCTTATGC	CGCTAACTTC
151	CCCGTTTGGG	CGGATCAGGC	AAACGCGATG	GTGCAGTATG	CCGTTTGGAC
201	GACACTTGCC	GCGGTCGGCG	TAGGTGCAA	CCTGCAACAT	TACAATCCCT
251	TGCCCGATGC	GGCGATTGCC	AAAGCGTGGA	ATATCCCCGA	AAACTGGTTG
301	TTGCGCGCAG	AAATGGTTAT	CGGCGGTATT	GAAGGGGCGG	CAGGTGAAAA
351	GACCTTTGAA	CCCGTTGCAG	AACGTTTGAA	AGTGTTCCGC	GCATAA

This corresponds to the amino acid sequence <SEQ ID 658; ORF6-1>:

```

45      1  ..LRVVVPADSF EPTAQKLNLF KAGAAATILFY EDQNVVKGLQ EQFPAYAAANF
      51  PVWADQANAM VQYAVWTTLA AVGVGANLQH YNPLPDAAIA KAWNIPENWL
     101  LRAOMVIGGI EGAAGEKTFF PVAERLKVFG A*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF6 shows 98.6% identity over a 140aa overlap with an ORF (ORF6a) from strain A of *N. meningitidis*:

-377-

```

      orf6.pep                                KVVQFVEXPLRAVVPADSFEPTAQKLNLFK
      orf6a                                QIVEHAVLHTPSSFNSQSARVVVLFGEEDKVVQFVEDALRAVVPADSFEPTAQKLNLFK
5      40          50          60          70          80          90

      orf6.pep                                AGAATILFYEDQNVVKGLQEQQPAYAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHY
      orf6a                                AGAATILFYEDQNVVKGLQEQQPAYAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHY
10     100        110        120        130        140        150

      orf6.pep                                NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX
      orf6a                                NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX
15     160        170        180        190        200

```

The complete length ORF6a nucleotide sequence <SEQ ID 659> is:

```

1  ATGACCCGTC AATCTCTGCA ACAGGCTGCC GAAAGCCGCC GTTCCATTTA
20 51  TTCGTAAAT  AAAAATCTGC CCGTCGGCAA AGATGAAATC GTCCAAATCG
   101  TCGAACACGC CGTTTTCAC ACACCTTCTT CGTTCAATTC CCAATCTGCC
   151  CGTGTGGTCG TCGTGTGTTG CGAAGAGCAT GATAAGGTGT GGCAATTTGT
   201  CGAAGACGCG CTGCGTGCCG TCGTGCCTGC CGACAGTTT  GAACCGACCG
   251  CGCAAAAATT GAACCTGTTT AAGGCGGGTG CGGCAACTAT TTTGTTTTAT
   301  GAAGATCAAA ATGTCGTCAA AGGTTTGAG GAGCAGTTC  CTGCTTATGC
   351  CGCCAACTTT CCCGTTTGGG CGGACCAGGC GAACGCGATG GTGCAGTATG
   401  CCGTTTGGAC GACACTTGCC GCGGTCGGCG TAGGTGCAAA CCTGCAACAT
   451  TACAATCCCT TGCCCGATGC GCGGATTGCC AAAGCGTGGA ATATCCCCGA
   501  AAACTGGTTG TTGCGCGCAC AAATGGTTAT CGGCGGTATT GAAGGGGCGG
   551  CAGGTGAAAA GACCTTTGAA CCAGTTGCAG AACGTTTGAA AGTGTTCCGGC
30 601  GCATAA

```

This is predicted to encode a protein having amino acid sequence <SEQ ID 660>:

```

1  MTRQSLQQA ESRRSIYSLN KNLVPGKDEI VQIVEHAVLH TPSSFNSQSA
35 51  RVVVLFGEEH DKVVQFVEDA LRAVVPADSF EPTAQKLNLF KAGAATILFY
   101  EDQNVVKGLQ EQFPAYAANF PVWADQANAM VQYAVWTTLA AVGVGANLQH
   151  YNPLPDAAIA KAWNIPENWL LRAQMVIIGI EGAAGEKTFE PVAERLKVFG
201  A*

```

ORF6a and ORF6-1 show 100.0% identity in 131 aa overlap:

```

40     orf6a.pep      50          60          70          80          90         100
      TPSSFNSQSARVVVLFGEEDKVVQFVEDALRAVVPADSFEPTAQKLNLFKAGAATILFY
      orf6-1                                LRAVVPADSFEPTAQKLNLFKAGAATILFY
                                         10          20          30

45     orf6a.pep      110         120         130         140         150         160
      EDQNVVKGLQEQQPAYAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHYNPLPDAAIA
      orf6-1          EDQNVVKGLQEQQPAYAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHYNPLPDAAIA
                                         40          50          60          70          80          90

50     orf6a.pep      170         180         190         200
      KAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX
      orf6-1          KAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX
                                         100         110         120         130

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF6 shows 95.7% identity over a 140aa overlap with a predicted ORF (ORF6ng) from *N.gonorrhoeae*:

-378-

	orf6.pep	KVWQFVEXPLRAVVPADSFEPTAQKLNLFK	30
	orf6ng	SNVSLDMSNPTVLRMGLPLYIASLRRGAIYKVWQFVEDALRAVVPADSFEPTAQKLFK	64
5	orf6.pep	AGAATILFYEDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHY	90
	orf6ng	AGAATILFYEDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHY	124
10	orf6.pep	NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGA	140
	orf6ng	NPLPDVAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKVFFPVAERLKVFGA	174

The complete length ORF6ng nucleotide sequence <SEQ ID 661> was identified as:

15	1	ATG	GCCGTTG	CGT	CAATGT	CAG	CTGGAT	ATG	TCCAATC	CTA	CGGTGTT
	51	ACG	CATGGGA	TTA	CCCTTAT	AT	TGCGTC	CCT	AAGAAGG	GGC	GCAATAT
	101	ATA	AGGTGTG	GCA	ATTTGTC	GA	AGACGCG	TGC	GTGCCGT	CGT	GCCTGCC
	151	GAC	AGTTTG	AAC	CGACCGC	GCA	AAAATG	AAG	CTGTTA	AGG	CGGGCGC
	201	GGC	AACCAT	TTG	TTTTATG	AAG	ATCAAAA	TGT	CGTCAA	GGT	TTGCAGG
	251	AGC	AGTTCCC	TGT	TATGCC	GCC	AACTTC	CCG	TTGGGC	GG	ACCAGCG
20	301	AAC	GCTATG	TAC	AGTATG	CGT	CTGGAC	ACA	CTGCCG	CGG	TCGGTGC
	351	AGG	TGCAAT	CTG	CAACATT	ACA	ACCCCTT	GCC	CGATGTG	GCG	ATTGCTA
	401	AAG	CGTGGA	TAT	TCCCGAA	AA	CTGGCTGT	TGC	GCGCGCA	AAT	GGTTATC
	451	GGT	GGTATTG	AAG	GGGcggc	aggt	gaaaaa	gtc	tttgaa	CCG	TGcgga
	501	acg	tttGAAA	GTG	TTCCGCG	CATAA					

25 This encodes a protein having amino acid sequence <SEQ ID 662>:

30	1	MAVASNVSLD	MSNPTVLRMG	LPLYIASLRR	GAIYKVWQFV	EDALRAVVA
	51	DSFEPTAQKL	KLFKAGAATI	LFYEDQNVVK	GLQEQFPAYA	ANFPVWADQA
	101	NAMVQYAVWT	TAAVGVGAN	LQHYNPLPDV	AIKAWNIP	NWLLRAQMVI
	151	GGIEGAAGEK	VFEPVAERLK	VFGA*		

ORF6ng and ORF6-1 show 96.9% identity in 131 aa overlap:

	orf6-1.pep				10	20	30
					LRAVVPADSFEPTAQKLNLFK	KAGAATILFY	
35	orf6ng	PTVLRMGLPLYIASLRRGAIYKVWQFVEDALRAVVPADSFEPTAQKLFK					
		20	30	40	50	60	70
	orf6-1.pep						
40		40	50	60	70	80	90
	orf6ng	EDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHYNPLPDAAIA					
		80	90	100	110	120	130
	orf6-1.pep						
45		100	110	120	130		
	orf6ng	KAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX					
		140	150	160	170		

50 It is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 79

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 663>

55	1	..GGCTACA	ACT	ACCTGTT	CGC	GCGCGGC	CAGC	CGCATCG	CCA	ACTACCA	AAAT
	51	CAACGGC	CATC	CCCGTTG	CCG	ACGCGCT	GGC	CGATACG	GGG	CAATGCC	AAC
	101	ACCGCCG	CCT	ATGAGCG	CGT	AGAAGTC	GTG	CGCGGCG	TGG	CGGGGCT	GCT
	151	GGACGGC	CAG	GGCGAGC	CCTT	CGCCACC	CGT	CAATCTG	GGT	CGCAAAC	GCC
	201	TGACCCG	CAA	GCCATTG	TTT	GAAGTCC	GCG	CCGAAGC	gGG	CAACCGc	AAA

251 CATTTCGGGC TGGACGCGGA CGTATCGGGC AGCCTGAACA CCGAAG.crC  
 301 rCTGCGCGGC CGCCTGGTTT CCAcCTTCGG ACGCGGCGAC TCGTGGCGGC  
 351 GGCGCGAAGC CAGCCGsKaT GCCGAACCTCT ACGGCATTTT GGAATACGAC  
 401 ATCGCACCGC AAACCCGCGT CCACGCAGrGC ATGGACTACC AGCAGGCGAA  
 5 451 AGAAACCGCC GACGCGCCGC TCAGcTACGC CGTGTACGAC AGCCAAGGTT  
 501 ATGCCACCGC CTTGCGCCCG AAAGACAACC CCGCCACAAA TTGGGCGAAC  
 551 AGCCACCACC GTGCGCTCAA CCTGTTCCGC GGCATCGAAC ACCGCTTCAA  
 601 CCAAGACTGG AAACCTCAAAG CCGAATACGA CTAC..

This corresponds to the amino acid sequence <SEQ ID 664; ORF23>:

10 1 ..GYNLYFARGS RIANYQINGI PVADALADTG NANTAAYERV EVVRGVAGLL  
 51 DGTGEPsATV NLVRKRLTRK PLFEVRAEAG NRKHFGLDAD VSGSLNTEXX  
 101 LRGRVLSTFG RGDSWRRRER SRXAELYGIL EYDIAPQTRV HAXMDYQQAK  
 151 ETADAPLSYA VYDSQGYATA FGPKDNPATN WANSHHRLN LFAGIEHFRN  
 201 QDWKLKAEYD Y..

15 Further work revealed the complete nucleotide sequence <SEQ ID 665>:

1 ATGACACGCT TCAAATATTC CCTGCTGTTT GCCGCCCTGT TGCCCGTGTA  
 51 CGCGCAGGCC GATGTTTCTG TTTCAGACGA CCCCAAACCG CAGGAAAGCA  
 101 CTGAATTGCC GACCATCACC GTTACCGCCG ACCGCACCGC GAGTTCCAAC  
 20 151 GACGGCTACA CTGTTTCCGG CACGCACACC CCGCTCGGGC TGCCCATGAC  
 201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC  
 251 GCGACCAAAA CATCAAAACG CTCGACCGCG CCCTGTTGCA GGCGACCGGC  
 301 ACCAGCCGCC AGATTTCGG CTCCGACCGC GCGGGCTACA ACTACCTGTT  
 351 CGCGCGCGGC AGCCGCATCG CCAACTACCA AATCAACGGC ATCCCCGTTG  
 401 CCGACGCGCT GGCGGATACG GGCAATGCCA ACACCGCCGC CTATGAGCGC  
 25 451 GTAGAAGTCG TCGCGGCGGT GCGGGGGCTG CTGGACGGCA CGGGCGAGCC  
 501 TTCCGCCACC GTCAATCTGG TGCGCAAACG CCTGACCCGC AAGCCATTGT  
 551 TTGAAGTCCG CGCCGAAGCG GGCAACCGCA AACATTTCGG GCTGGACGCG  
 601 GACGTATCGG GCGCCTGAA CACCGAAGGC ACGCTGCGCG GCCGCTGGT  
 651 TTCCACCTTC GGACGCGGCG ACTCGTGCGG GCGGCGCGAA CGCAGCCGCG  
 30 701 ATGCCGAACCT CTACGGCATT TTGGAATACG ACATCGCACC GCAAACCCGC  
 751 GTCCACGCAG GCATGGACTA CCAGCAGGCG AAAGAAACCG CCGACGCGCC  
 801 GCTCAGCTAC GCCGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGGCC  
 851 CGAAAGACAA CCCCGCCACA AATTGGGCGA ACAGCCGCCA CCGTGCGCTC  
 901 AACCTGTTCG CCGGCATCGA ACACCGCTTC AACCAAGACT GGAAACTCAA  
 35 951 AGCCGAATAC GACTACACCC GCAGCCGCTT CCGCCAGCCC TACGGCGTAG  
 1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CCTGATTCCC  
 1051 GGTATTGGC ACGCCGACCC GCGCACCCAC AGCGCCAGCG TGTCAATTGAT  
 1101 CGGCAAATAC CGCTGTTCG GCCGGAACA CGATTTAATC GCGGGTATCA  
 40 1151 ACGGTTACAA ATACGCCAGC AACAAATACG GCGAACGCAG CATCATCCCC  
 1201 AACGCCATT CCAACGCCTA CGAATTTTCC CGCACGGGTG CCTACCCGCA  
 1251 GCCTGCATCG TTTGCCCAA CCATCCCGCA ATACGGCACC AGGCGGCAAA  
 1301 TCGGCGGCTA TCTCGCCACC CGTTTCCGCG CCGCCGACAA CCTTTCGCTG  
 1351 ATTTTGGGCG GACGATACAC CCGTTACCGC ACCGGCAGCT ACGACAGCCG  
 45 1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTCACC CCCTACACAG  
 1451 GCATCGTGTT CGACCTGACC GGCAACCTGT CTCTTTACGG CTCGTACAGC  
 1501 AGCCTGTTCG TCCCGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA  
 1551 ACCCGTAACC GGCAACAATC TGAAGCCGG CATCAAAGGC GAATGGCTTG  
 1601 AAGGCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC  
 1651 CTCGCCACCG CAGCAGGACG CGACCCGAGC GGCAACACCT ACTACCGCGC  
 50 1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GGCGGCCGCA  
 1751 TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCGC  
 1801 GACCAAGACG GCAGCCGCCT GAACCCCGAC AGCGTACCCG AACGCAGCTT  
 1851 CAAACTCTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCC AGCGGCTGGA  
 55 1901 CCATCGGCGC AGGCGTGCGC TGGCAGAGCG AAACCCACAC CGACCCTGCC  
 1951 ACGCTCCGCA TCCCCAACCC CGCCGCCAAA GCCCGCGCCG CCGACAACAG  
 2001 CCGCCAAAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA  
 2051 ATCCGCGCGC CGAACTGTCG CTGAACGTGG ACAATCTGTT CAACAAACAC  
 2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GGCGCACTGC GGACAGTGAA  
 2151 CGCGGCGTTT ACCTATCGGT TTAAATAA

60 This corresponds to the amino acid sequence <SEQ ID 666; ORF23-1>:

1 MTRFKYSLLE AALLPVYAQA DVSVDPPKP QESTELPTIT VTADRTASSN  
 51 DGYTVSGTHT PLGLPMTLRE IPQSVSVITS QQMRDQNIKT LDRALLQATG  
 101 TSRQIYGSDR AGYNLYFARG SRIANYQING IPVADALADT GNANTAAYER  
 151 VEVVRGVAGL LDGTGEPsAT VNLVRKRLTR KPLFEVRAEA GNRKHFGLDA  
 65 201 DVSGSLNTEG TLRGRVLSTF GRGDSWRRRE RSRDAELYGI LEYDIAPQTR

-380-

251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL  
 301 NLFAGIEHRF NQDWKLKAEY DYTRSFRFQP YGVAGVLSID HNTAATDLIP  
 351 GYWHADPRTH SASVSLIGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP  
 401 NAIPNAYEFS RTGAYPQPAS FAQTIPQYGT RRQIGGYLAT RFRAADNLSL  
 5 451 ILGGRYTRYR TGSYDSRTQG MTYVSANRET PYTGIVFDLT GNLSLYGSYS  
 501 SLFVPQSQKD EHGSYLKPVV GNNLEAGIKG EWLEGRNLAS AAVYRARKNN  
 551 LATAAGRDPS GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR  
 601 DQDGSRLNPD SVPERSFKLF TAYHFAPEAP SGWTIGAGVR WQSEHTDPA  
 651 TLRIPNPAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVNLFNKH  
 10 701 YRTQPDHRSY GALRTVNAAF TYRFK\*

Computer analysis of this amino acid sequence gave the following results:

Homology with the ferric-pseudobactin receptor PupB of *Pseudomonas putida* (accession number P38047)

ORF23 and PupB protein show 32% aa identity in 205aa overlap:

15 Orf23 6 FARGSRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRK 65  
 ++RG I NY+++G+P + L D + + A ++RVE+VRG GL+ G G PSAT+NL+RK  
 PupB 215 WSRGFAIQNYEVDGVPTSTRL-DNYSQSMAMFDRVEIVRGATGLISGMGNPSATINLIRK 273  
 Orf23 66 RLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGRLVSTFXXXXXXXXXXXXXAE 125  
 R T + + EAGN +G DVSG L +RGR V+ +  
 20 PupB 274 RPTAEAQASITGEAGNWDRYGTGFDVSGPLTETGNIRGRFVADYKTEKAWIDRYNQSQL 333  
 Orf23 126 LYGILEYDIAPQTRVHAXMDYQQA KETADAPLSYAVYD--SQGYATAFGPKDNPATNWAN 183  
 +YGI E+D++ T + Y + D+PL + S G T N A +W+  
 25 PupB 334 MYGITEFDLSEDTLLTVGFSY--LRSDIDSPLRSGLPTRFSTGERTNLKRSLNAAPDWSY 391  
 Orf23 184 SHHRALNLFAGIEHRFNQDWKLKAE 208  
 + H + F IE + W K E  
 PupB 392 NDHEQTSFFTSIEQQLGNGWSGKIE 416

30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF23 shows 95.7% identity over a 211aa overlap with an ORF (ORF23a) from strain A of *N.*

*meningitidis*:

35 orf23.pep GYNLYFARGSRIANYQINGIPVADALADTG  
 orf23a QMRDQNIKALDRALLQATGTSRQIYGSDRAGYNLYFARGSRIANYQINGIPVADALADTG  
 90 100 110 120 130 140  
 40 orf23.pep NANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDAD  
 orf23a NANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRPTRKPLFEVRAEAGNRKHFGLDAD  
 150 160 170 180 190 200  
 45 orf23.pep VSGSLNTEXXLRGRLVSTFGRGDSWRRRERSRXAELYGILEYDIAPQTRVHAXMDYQQA K  
 orf23a VSGSLNAEGTLRGRLVSTFGRGDSWRQRERSRDAELYGILEYDIAPQTRVHAGMDYQQA K  
 210 220 230 240 250 260  
 50 orf23.pep ETADAPLSYAVYDSQGYATAFGPKDNPATNWANSHHRALNLFAGIEHRFNQDWKLKAEYD  
 orf23a ETADAPLSYAVYDSQGYATAFGPKDNPATNWANSHHRALNLFAGIEHRFNQDWKLKAEYD  
 270 280 290 300 310 320  
 60 orf23.pep Y  
 orf23a YTRSFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTHSASVSLIGKYRLFGREHDLIA  
 330 340 350 360 370 380

The complete length ORF23a nucleotide sequence <SEQ ID 667> is:

-381-

1 ATGACACGCT TCAAATATTC CCTGCTGTTT GCCGCCCTGT TGCCCCGTGTA  
 51 CGCGCAGGCC GATGTTTCTG TTTCAGACGA CCCCCAACCG CAGGAAAGCA  
 101 CTGAATTGCC GACCATCACC GTTACCGCCG ACCGCACCGC GAGTTCCAAC  
 151 GACGGCTACA CTGTTTCCGG CACGCACACC CCGCTCGGGC TGCCCATGAC  
 201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC  
 251 GCGACCAAAA CATCAAAGCG CTCGACCGCG CCCTGTTGCA GGCGACCGGC  
 301 ACCAGCCGCC AGATTTACGG CTCCGACCGC GCGGGCTACA ACTACCTGTT  
 351 CGCGCGCGGC AGCCGCATCG CCAACTACCA AATCAACGGC ATCCCCGTG  
 401 CCGACGCGCT GGCCGATACG GGCAATGCCA ACACCGCCGC CTATGAGCGC  
 451 GTAGAAGTCG TGC CGCGGCTG CTGGACGGCA CCGGCGAGCC  
 501 TTCCGCCACC GTCAATCTGG TGC GCAAACG CCCGACCGC AAGCCATTGT  
 551 TTGAAGTCCG CGCCGAAGCG GGCAACGCCA AACATTTCGG GCTGGCGCGC  
 601 GACGTATCGG GCAGCCTGAA TGCCGAAGGC ACGCTGCGCG GCCGCTGGT  
 651 TTCCACCTTC GGACGCGCGG ACTCGTGGCG GCAGCGCGAA CGCAGCCGCG  
 701 ATGCCGAAC CTACGGCATT TTGGAATACG ACATCGCACC GCAAACCCGC  
 751 GTCCACGCAG GCATGGACTA CCAGCAGGCG AAAGAAACCG CCGACGCGCC  
 801 GCTCAGCTAC GCCGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGGCC  
 851 CGAAAGACAA CCCC GCCACA AATTGGGCGA ACAGCCGCCA CCGTGCCTC  
 901 AACCTGTTTC CCGGCATCGA ACACCGCTTC AACCAAGACT GGAAACTCAA  
 951 AGCCGAATAC GACTACACCC GCAGCCGCTT CCGCCAGCCC TACGGCGTAG  
 1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CCTGATTCCC  
 1051 GGTATTGGC ACGCCGACCC GCGCACCAC AGCGCCAGCG TGTCATTAAT  
 1101 CGGCAATAC CGCCTGTTTC GCGCGAACA CGATTTAATC GCGGGTATCA  
 1151 ACGGTTACAA ATACGCCAGC AACAAATACG GCGAACGCGC CATCATCCCC  
 1201 AACGCCATT CCAACGCCTA CGAATTTTCC CGCACGGGTG CCTACCCGCA  
 1251 GCCTGCATCG TTTGCCCAA CCATCCCGCA ATACGGCACC AGGCGGCAAA  
 1301 TCGGCGGCTA TCTCGCCACC CGTTTCCGCG CCGCCGACAA CCTTTCGCTG  
 1351 ATACTCGGCG GCAGATACAG CCGTTACCGC ACCGGCAGCT ACGACAGCCG  
 1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTCACC CCTACACAG  
 1451 GCATCGTGTT CGACCTGACC GGCAACCTGT CGCTTTACGG CTCGTACAGC  
 1501 AGCCTGTTTC TCCCGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA  
 1551 ACCCGTAACC GGCAACAATC TGGAAGCCGG CATCAAAGGC GAATGGCTTG  
 1601 AAGGCCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC  
 1651 CTCGCCACCG CAGCAGGACG CGACCCGAGC GGCAACACCT ACTACCGCGC  
 1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GGCGGCCGCA  
 1751 TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCGC  
 1801 GACCAAGACG GCAGCCGCTT GAACCCCGAC AGCGTACCCG AACGCAGCTT  
 1851 CAAACTCTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCCC AGCGGTGGA  
 1901 CCATCGGCGC AGGCGTGCCT TGGCAGAGCG AAACCCACAC CGACCTGCC  
 1951 ACGCTCCGCA TCCCAACCC CGCCGCCAAA GCCCGCGCCG CCGACAACAG  
 2001 CCGCCAAAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA  
 2051 ATCCGCGCGC CGAACTGTCT CTGAACGTGG ACAATCTGTT CAACAACAC  
 2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GGCGCACTGC GGACAGTGAA  
 2151 CGCGGCGTTT ACCTATCGGT TTAAATAA

45 This encodes a protein having amino acid sequence <SEQ ID 668>:

1 MTRFKYSLLF AALLPVYAQA DVSVSDPKP QESTELPTIT VTADRTASSN  
 51 DGYTVSGTHT PLGLPMTLRE IPQSVSVITS QOMRDQNIKA LDRALLQATG  
 101 TSRQIYGS DR AGNYL FARG SRIANYQING IPVADALADT GNANTAYER  
 151 VEVVRGVAGL LDGTGEPSAT VNLVRKRPT R KPLFEVRAEA GNRKHFGLGA  
 201 DVSGSLNAEG TLRGRLVSTF GRGDSWRQRE RSRDAELYGI LEYDIAPQTR  
 251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL  
 301 NLFAGIEHRF NQDWKLKAEY DYTRSFRFP YGVAGVLSID HNTAATDLIP  
 351 GYWHADPRTH SASVSLIGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP  
 401 NAIPNAYEFS RTGAYPQPAS FAQTIPQYGT RRQIGGYLAT RFRAADNLSL  
 451 ILGGYRSRYR TGSYDSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS  
 501 SLFVPSQSKD EHGSLYKPV T GNNLEAGIKG EWLEGRNLAS AAVYRARKNN  
 551 LATAAGRDP S GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR  
 601 DQDGSRLNPD SVPERSFKLF TAYHFAPEAP SGWTIGAGVR WQSEHTDPA  
 651 TLRI PNPAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVNLFNKH  
 701 YRTQPRHSY GALRTVNAAF TYRFK\*

ORF23a and ORF23-1 show 99.2% identity in 725 aa overlap:

	10	20	30	40	50	60
orf23a.pep	MTRFKYSLLF AALLPVYAQA DVSVSDPKP QESTELPTIT VTADRTASSN DGYTVSGTHT					
orf23-1	MTRFKYSLLF AALLPVYAQA DVSVSDPKP QESTELPTIT VTADRTASSN DGYTVSGTHT					
	10	20	30	40	50	60

-382-

		70	80	90	100	110	120
	orf23a.pep	PLGLPMTLREIPQSVSVITSQQMRDQNIKALDRALLQATGTSRQIYGSDRAGYNYLFARG					
5	orf23-1	PLGLPMTLREIPQSVSVITSQQMRDQNIKTLDRAALLQATGTSRQIYGSDRAGYNYLFARG					
		70	80	90	100	110	120
	orf23a.pep	130	140	150	160	170	180
	orf23a.pep	SRIANYQINGIPVADALADTGNANTAAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRPRTR					
10	orf23-1	SRIANYQINGIPVADALADTGNANTAAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRLTR					
		130	140	150	160	170	180
	orf23a.pep	190	200	210	220	230	240
	orf23a.pep	KPLFEVRAEAGNRKHFG LGADVSGSLNAEGTLRGRLVSTFGRGDSWRQRERSRDAELYGI					
15	orf23-1	KPLFEVRAEAGNRKHFG LDADVSGSLNTEGTLRGRLVSTFGRGDSWRRRERSRDAELYGI					
		190	200	210	220	230	240
	orf23a.pep	250	260	270	280	290	300
	orf23a.pep	LEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRAL					
20	orf23-1	LEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRAL					
		250	260	270	280	290	300
	orf23a.pep	310	320	330	340	350	360
	orf23a.pep	NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHN TAATDLIPGYWHADPRTH					
25	orf23-1	NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHN TAATDLIPGYWHADPRTH					
		310	320	330	340	350	360
30	orf23a.pep	370	380	390	400	410	420
	orf23a.pep	SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS					
35	orf23-1	SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS					
		370	380	390	400	410	420
	orf23a.pep	430	440	450	460	470	480
	orf23a.pep	FAQTIPQYGTTRQIGGYLATRFRAADNLSLILGGRYSR YRTGSYDSRTQGMYTVSANRFT					
40	orf23-1	FAQTIPQYGTTRQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMYTVSANRFT					
		430	440	450	460	470	480
	orf23a.pep	490	500	510	520	530	540
	orf23a.pep	PYTGIVFDLTGNLSLYGSYSSLFV PQSQKDEHGSYLKPVTGNNLEAGIKGEWLEGRNLAS					
45	orf23-1	PYTGIVFDLTGNLSLYGSYSSLFV PQSQKDEHGSYLKPVTGNNLEAGIKGEWLEGRNLAS					
		490	500	510	520	530	540
	orf23a.pep	550	560	570	580	590	600
	orf23a.pep	AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR					
50	orf23-1	AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR					
		550	560	570	580	590	600
	orf23a.pep	610	620	630	640	650	660
	orf23a.pep	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSETH TDPATLRIPNPAAK					
55	orf23-1	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSETH TDPATLRIPNPAAK					
		610	620	630	640	650	660
60	orf23a.pep	670	680	690	700	710	720
	orf23a.pep	ARAADNSRQKAYAVADIMARYRFNPRAELSLNVDNLFNKHYRTQPD RHSYGALRTVNAAF					
65	orf23-1	ARAADNSRQKAYAVADIMARYRFNPRAELSLNVDNLFNKHYRTQPD RHSYGALRTVNAAF					
		670	680	690	700	710	720
	orf23a.pep	TYRFXK					
70	orf23-1	TYRFXK					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF23 shows 93.4% identity over a 211aa overlap with a predicted ORF (ORF23.ng) from *N. gonorrhoeae*:

5	orf23.pep	GYNYLFARGSR IANYQINGIPVADALADTGNANTAAAYERVEVVVRGVAGLLD	51
	orf23ng	SAVDACRIPGYNYLFARGSR IANYQINGIPVADALADTGNANTAAAYERVEVVVRGVAGLPD	60
10	orf23.pep	GTGEPSATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGLVSTFGR	111
	orf23ng	GTGEPSATVNLVRKHPTKPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGLVSTFGR	120
15	orf23.pep	GDSWRRRERSRXAELYGILEYDIAPQTRVHAXMDYQQAKETADAPLSYAVYDSQGYATAF	171
	orf23ng	GDSWRQLERSRDAELYGILEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAF	180
	orf23.pep	GPKDNPATNWNASHHRLNLFAGIEHRFNQDWKLKAEYDY	211
	orf23ng	GPKDNPATNWSNSRNRALNLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHS	240

The ORF23ng nucleotide sequence <SEQ ID 669> is predicted to encode a protein comprising  
20 amino acid sequence <SEQ ID 670>:

	1	SAVDACRIPG	YNYLFARGSR	IANYQINGIP	VADALADTGN	ANTAAYERVE
	51	VVRGVAGLPD	GTGEPSATVN	LVRKHPTKRP	LFEVRAEAGN	RKHFGLGADV
	101	SGSLNAEGLT	RGRVSTFGR	GDSWRQLERS	RDAELYGILE	YDIAPQTRVH
	151	AGMDYQQAKE	TADAPLSYAV	YDSQGYATAF	GPKDNPATNW	SNSRNRALNL
25	201	FAGIEHRFNQ	DWKLKAEYDY	TRSRFRQPYG	VAGVLSIDHS	TAATDLIPGY
	251	WHADPRTHSA	SMSLTGKYRL	FGREHDLIAG	INGYKYASNK	YGERSIIPNA
	301	IPNAYEFSRT	GAYPQPSSFA	QTIPQYDTRR	QIGGYLATRF	RAADNLSLIL
	351	GGYRSRYRAG	SYNSRTQGMT	YVSANRETPY	TGIVFDLTGN	LSLYGSYSSL
	401	FVPQLQKDEH	GSYLKPVTGN	NLEADIKGEW	LEGRNLASAA	VYRARKNNLA
30	451	TAAGRDQSGN	TYYRAANQAK	THGWEIEVGG	RITPEWQIQA	GYSQSKPRDQ
	501	DGSRLNPDSV	PERSFKLFTA	YHLAPEAPSG	RTIGAGVRRQ	GETHTDPAAL
	551	RIPNPAAKAR	AVANSRQKAY	AVADIMARYR	FNPRTELSLN	VDNLFNKHYR
	601	TQFDRHSYGA	LRTVNAAFTY	RFK*		

Further work revealed the complete nucleotide sequence <SEQ ID 671>:

35	1	ATGACACGCT	TCAAATACTC	CCTGCTTTT	GCCGCCCTGC	TACCCGTGTA
	51	CGCGCAGGCC	GATGTTTCTG	TTTCAGACGA	CCCCAAACCG	CAGGAAAGCA
	101	CCGAATTGCC	GACCATCACC	GTTACCGCCG	ACCGCACCGC	GAGTCCAAC
	151	GACGGCTACA	CCGTTTCCGG	CACGCACACC	CCGTTCGGGC	TGCCCATGAC
	201	CCTGCGCGAA	ATCCCGCAGA	GCGTCAGCGT	CATCACATCG	CAACAAATGC
40	251	GCGACCAAAA	CATCAAAACG	CTCGACCGCG	CCCTGTTGCA	GGCGACCGGC
	301	ACCAGCCGCC	AGATTTACGG	CTCCGACCGC	GCGGGCTACA	ACTACCTGTT
	351	CGCGCGCGGC	AGCCGCATCG	CCAACACCA	AATCAACGGC	ATCCCGCTTG
	401	CCGACGCGCT	GGCCGATACG	GGCAATGCCA	ACACCGCCGC	CTATGAGCGC
	451	GTAGAAGTCG	TGCGCGGCGT	GGCGGGGCTG	CCGGACGGCA	CGGGCGAGCC
45	501	TTCTGCCACC	GTCAATCTGG	TACGCAAACA	CCCGACCCGC	AAGCCATTGT
	551	TTGAAGTCCG	CGCCGAAGCC	GGCAACCGCA	AACATTTCCG	GCTGGGCGCG
	601	GACGTATCGG	GCAGCTGAA	CGCCGAAGGC	ACGCTGCGCG	GCCGCTTGGT
	651	TTCCACCTTC	GGACGCGGCG	ACTCGTGGCG	GCAGCTCGAA	CGCAGCCGCG
	701	ATGCCGAAC	CTACGGCATT	TTGGAATACG	ACATCGCACC	GCAAACCCGC
50	751	GTCCACGCAG	GCATGGACTA	CCAGCAGGCG	AAAGAAACCG	CAGACGCGCC
	801	GCTCAGCTAC	GCCGTGTACG	ACAGCCAAGG	TTATGCCACC	GCCTTCGGCC
	851	CAAAAGACAA	CCCCGCCACA	AATTGGTCTGA	ACAGCCGCAA	CCGTGCGCTC
	901	AACCTGTTCT	CCGGCATAGA	ACACCGCTTC	AACCAAGACT	GGAAACTCAA
	951	AGCCGAATAC	GACTACACCC	GTAGCCGCTT	CCGCCAGCCC	TACGGTGTGG
55	1001	CAGGCGTACT	TTCCATCGAC	CACAGCACTG	CCGCCACCGA	CCTGATTCCC
	1051	GGTTATTGGC	ACGCcgatcc	GCGCACCCAC	AGCGCCAGCA	TGTCATTGAC
	1101	CGGCAAATAC	CgcctGTTCTG	GCCGCGAGCA	CGATTTAATC	GCGGCTATCA
	1151	ACGGCTACAA	ATACGCCAGC	AACAAATACG	GCGAACGCAG	CATCATTTCC
	1201	AACGCCATT	CCAACGCCTA	CGAATTTTCC	CGCACGGGCG	CCTATCCGCA
60	1251	GCCATCATCG	TTTGCCCAAA	CCATCCCGCA	ATACGACACC	AGGCGGCAAA
	1301	TCGGCGGCTA	TCTCGCCACC	CGTTTCCGCG	CCGCCGACAA	CCTTTCGCTG
	1351	ATACTCGGCG	GCAGATACAG	CCGCTACCGC	GCAGGCAGCT	ACAACAGCCG



-384-

1401	CACACAAGGC	ATGACCTATG	TGTCCGCCAA	CCGTTTCACC	CCCTACACAG
1451	GCATCGTGTT	CGATCTGACC	GGCAACCTGT	CGCTTTACGG	CTCGTACAGC
1501	AGCCTGTTCG	TCCCGCAATT	GCAAAAAGAC	GAACACGGCA	GCTACCTGAA
1551	ACCCGTAACC	GGCAACAATC	TGGAAGCCGA	CATCAAAGGC	GAATGGCTTG
5	1601	AAGGGCGTCT	GAACGCATCC	GCCGCCGTGT	ACCGCGCCCG
	1651	CTCGCCACCG	CAGCAGGACG	CGACCAGAGC	GGCAACACCT
	1701	CGCCAACCAA	GCCAAAACCC	ACGGCTGGGA	AATCGAAGTC
	1751	TCACGCCCGA	ATGGCAGATA	CAGGCAGGCT	ACAGCCAAAG
	1801	GACCAAGACG	GCAGCCGCCT	GAACCCCGAC	AGCGTAACCG
10	1851	CAAACCTCTT	ACCGCCTACC	ACTTAGCCCC	CGAAGCCCCC
	1901	CCATcggtGC	GGGTGTGCGC	CGGCAGGGCG	AAACCCACAC
	1951	GCGCTCCGCA	TCCCCAACCC	CGCCGCCAAA	GCCCGCGCCG
	2001	CCGCCAGAAA	GCCTACGCCG	TCGCCGACAT	CATGGCGCGT
	2051	ATCCGCGCAC	CGAACTGTCT	CTGAACGTGG	ACAACCTGTT
15	2101	TACCGCACCC	AGCCCGACCG	CCACAGCTAC	GGCGCACTGC
	2151	CGCGGCGTTT	ACCTATCGGT	TTAAATAA	

This corresponds to the amino acid sequence <SEQ ID 672; ORF23ng-1>:

	1	MTRFKYSLLF	AALLPVYAQA	DVSVSDDPKP	QESTELPTIT	VTADRTASSN
	51	DGYTVSGTHT	PFGLPMTLRE	IPQSVSVITS	QQMRDQNIKT	LDRALLQATG
20	101	TSRQIYGSDR	AGYNILFARG	SRIANYQING	IPVADALADT	GNANTAAYER
	151	VEVVRGVAGL	PDGTGEPSAT	VNLVRKHPTR	KPLFEVRAEA	GNRKHFGLGA
	201	DVSGSLNAEG	TLRGLVSTF	GRGDSWRQLE	RSRDAELYGI	LEYDIAPQTR
	251	VHAGMDYQQA	KETADAPLSY	AVYDSQGYAT	AFGPKDNPAT	NWSNSRNRAL
25	301	NLFAGIEHRF	NQDWKLKAEY	DYTRSFRQPY	GVAGVLSIDH	NTAATDLIPG
	351	GYWHADPRTH	SASMSLTGKY	RLFGREHDLI	AGINGYKYAS	NKYGERSIIP
	401	NAIPNAYEFS	RTGAYPOPSS	FAQTIPQYDT	RRQIGGYLAT	RFRAADNLSL
	451	ILGGYRSRYR	AGSYNSRTQG	MTYVSANRFT	PYTGIVFDLT	GNLSLYGSYS
	501	SLFVPQLQKD	EHGSYLKPV	T		
30	551	LATAAGRDQS	GNTYYRAANQ	AKTHGWEIEV	GGRITPEWQI	QAGYSQSKPR
	601	DQDGSRLNPD	SVPERSFKLF	TAYHLAPEAP	SGRTIGAGVR	RQGETHTDPA
	651	ALRIPNPAAK	ARAVANSRQK	AYAVADIMAR	YRENPRTELS	LNVDNLFNKH
	701	YRTQPDHRSY	GALRTVNAAF	TYRFK*		

ORF23ng-1 and ORF23-1 show 95.9% identity in 725 aa overlap:

		10	20	30	40	50	60
35	orf23-1.pep	MTRFKYSLLF	AALLPVYAQA	DVSVSDDPKP	QESTELPTIT	VTADRTASSN	DGYTVSGTHT
	orf23ng-1	MTRFKYSLLF	AALLPVYAQA	DVSVSDDPKP	QESTELPTIT	VTADRTASSN	DGYTVSGTHT
		10	20	30	40	50	60
40	orf23-1.pep	PLGLPMTLRE	IPQSVSVITS	QQMRDQNIKT	LDRALLQATG	TSRQIYGSDR	AGYNILFARG
	orf23ng-1	PFGLPMTLRE	IPQSVSVITS	QQMRDQNIKT	LDRALLQATG	TSRQIYGSDR	AGYNILFARG
		70	80	90	100	110	120
45	orf23-1.pep	SRIANYQING	IPVADALADT	GNANTAAYER	VEVVRGVAGL	PDGTGEPSAT	VNLVRKRLTR
	orf23ng-1	SRIANYQING	IPVADALADT	GNANTAAYER	VEVVRGVAGL	PDGTGEPSAT	VNLVRKHPTR
		130	140	150	160	170	180
50	orf23-1.pep	KPLFEVRAEA	GNRKHFGLDA	DVSGSLNTEG	TLRGLVSTF	GRGDSWRRRR	ERSRDAELYGI
	orf23ng-1	KPLFEVRAEA	GNRKHFGLDA	DVSGSLNAEG	TLRGLVSTF	GRGDSWRQLER	SRDAELYGI
		190	200	210	220	230	240
55	orf23-1.pep	LEYDIAPQTR	VHAGMDYQQA	KETADAPLSY	AVYDSQGYAT	AFGPKDNPAT	NWANSRNRAL
	orf23ng-1	LEYDIAPQTR	VHAGMDYQQA	KETADAPLSY	AVYDSQGYAT	AFGPKDNPAT	NWSNSRNRAL
		250	260	270	280	290	300
60	orf23-1.pep	NLFAGIEHRF	NQDWKLKAEY	DYTRSFRQPY	GVAGVLSIDH	NTAATDLIPG	YWHADPRTH
	orf23ng-1	NLFAGIEHRF	NQDWKLKAEY	DYTRSFRQPY	GVAGVLSIDH	STAATDLIPG	YWHADPRTH
		310	320	330	340	350	360
65	orf23-1.pep	NLFAGIEHRF	NQDWKLKAEY	DYTRSFRQPY	GVAGVLSIDH	NTAATDLIPG	YWHADPRTH
	orf23ng-1	NLFAGIEHRF	NQDWKLKAEY	DYTRSFRQPY	GVAGVLSIDH	STAATDLIPG	YWHADPRTH

-385-

		310	320	330	340	350	360
		370	380	390	400	410	420
5	orf23-1.pep	SASVSLIGKYRLFGRHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS					
	orf23ng-1	SASMSLTGKYRLFGRHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPSS					
		370	380	390	400	410	420
10	orf23-1.pep	FAQTIPQYGTTRQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMYVSANRFT					
	orf23ng-1	FAQTIPQYDTRRQIGGYLATRFRAADNLSLILGGRYSTRYRAGSYNSRTQGMYVSANRFT					
		430	440	450	460	470	480
15	orf23-1.pep	PYTGIVFDLTGNLSLYGSYSSLFVPPSQKDEHGSYLKPVTGNNLEAGIKGEWLEGRNLAS					
	orf23ng-1	PYTGIVFDLTGNLSLYGSYSSLFVPPQLQKDEHGSYLKPVTGNNLEADIKGEWLEGRNLAS					
		490	500	510	520	530	540
20	orf23-1.pep	AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR					
	orf23ng-1	AAVYRARKNNLATAAGRDQSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKPR					
25		550	560	570	580	590	600
	orf23-1.pep	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSEHTDTPALRIPNPAK					
	orf23ng-1	DQDGSRLNPDSVPERSFKLFTAYHLAPEAPSGRTIGAGVRRQGETHTDPAALRIPNPAK					
30		610	620	630	640	650	660
	orf23-1.pep	ARAADNSRQKAYAVADIMARYRFPRAELSLNVDNLFNKHYRTQPDHRSYGALRTVNAAF					
	orf23ng-1	ARAVANSRQKAYAVADIMARYRFPRTLSLNVDNLFNKHYRTQPDHRSYGALRTVNAAF					
35		670	680	690	700	710	720
	orf23-1.pep	TYRFXK					
	orf23ng-1	TYRFXK					

In addition, ORF23ng-1 shows significant homology with an OMP from *E.coli*:

45	sp P16869 FHUE_ECOLI OUTER-MEMBRANE RECEPTOR FOR FE(III)-COPROGEN, FE(III)-FERRIOXAMINE B AND FE(III)-RHODOTRULIC ACID PRECURSOR >gi 1651542 gnl PID d1015403 (D90745) Outer membrane protein FhuE precursor [Escherichia coli]
50	>gi 1651545 gnl PID d1015405 (D90746) Outer membrane protein FhuE precursor [Escherichia coli] >gi 1787344 (AE000210) outer-membrane receptor for Fe(III)-coprogen, Fe(III)-ferrioxamine B and Fe(III)-rhodotrulic acid precursor [Escherichia coli] Length = 729 Score = 332 bits (843), Expect = 3e-90 Identities = 228/717 (31%), Positives = 350/717 (48%), Gaps = 60/717 (8%)
55	Query: 38 TITVTADRTASSN--DGYTVSGTHTPFGLPMTLREIPQSVSVITSQQMRDQNIKTLDRL 95 T+ V TA + + Y+V+ T + MT R+IPQSV+++ Q+M DQ ++TL + Sbjct: 43 TVIVEGSATAPDDGENDYSVTSTAGTKMQMTQRDIQSVTIVSQQRMEDQQQLQTLGEVM 102
60	Query: 96 LQATGTSRQIYGSDRAGYNLYFARGSRANYQINGIP-----VADALADTGNANTAA 147 G S+ SDRY ++RG +I NY ++GIP + DAL+D A Sbjct: 103 ENTLGISKSQADSRLY---YSRGFQIDNYMVDGIPTYFESRWNLGDALSDM-----AL 154
65	Query: 148 YERVEVVRGVAGLPDGTGEP SATVNLVRKHPTKPLF-EVRAEAGNRKHFLGADVSGSL 206 +ERVEVVRG GL GTG PSA +N+VRKH T + +V AE G+ AD+ L Sbjct: 155 FERVEVVRGATGLMTGTGNPSAAINMVRKHATSREFKGDVSAEYGSWNKERYVADLQSPL 214
70	Query: 207 NAEGTLRGRVSTFGRGDSWRQLERSRDAELYGILEYDIAPQTRVHAGMDYQOAKETADA 266 +G +R R+V + DSW S GI++ D+ T + AG +YQ+ + Sbjct: 215 TEDGKIRARIVGGYQNNDSWLDRYNSEKTFESGIVDADLGLTTLSAGYEYQRIDVNSPT 274
	Query: 267 PLSYAVYDSQGYATAFGPKDNPATNWSNSRNRLNLFAGIEHRFNQDWKLKAEYDYTRSR 326

```

          +++ G + ++      + A +W+ +      +F ++ +F W+      ++
Sbjct: 275 WGGLPRWNTDGSSNSYDRARSTAPDWAYNDKEINKVFMTLKQQFADTWQATLNATHSEVE 334
Query: 327 F--RQPYGVAGVLSIDHSTAA--TDLIPGY-----WHADPRTHSA-SMSLTGKYRLFG 374
5      F + Y A V D      ++ PG+      W++ R A + G Y LFG
Sbjct: 335 FDSKMMYVDAYVNKADGMLVGPYSNYGPGFDYVGGTGWNSGKRKVDALDLFADGSYELEFG 394
Query: 375 REHDLIAGINGYKYASNKYGER--SIIPNAIPNAYEFSRTGAYPQPSSFAQTIPQYDTRR 432
10      R+H+L+ G Y +N+Y +I P+ I + Y F+ G +PQ Q++ Q DT
Sbjct: 395 RQHNLMFG-GSYSKQNNRYFSSWANIFPDEIGSFYNFN--GNFPQTDWSPQSLAQDDTTH 451
Query: 433 QIGGYLATRFRAADNLSLILGGRYSRYRAGSYNSRTQGMTY-VSANRFTPYTGIVFDXXX 491
15      Y ATR AD L LILG RY+ +R + +TY + N TPY G+VFD
Sbjct: 452 MKSLYAATRVTLADPLHLILGARYTNWRVDT-----LTYSMEKNHTTPYAGLVFDIND 504
Query: 492 XXXXXXXXXXXXFPQLQKDEHGSYLKPVGTGNNLEADIKGEWLEGRNLASAAVYRARKNNL 551
20      F PQ +D G YL P+TGNN E +K +W+ RL + A++R ++N+
Sbjct: 505 NWSTYASYTSIFQPQNDRDSSGKYLAIPITGNNYELGLKSDWMNSRLTTTLAIFRIEQDNV 564
Query: 552 ATAAGR---DQSGNTYYRAANQAKTHGWEIEVGGRTPEWQIQAGYSQSKPRDQDGSRLN 608
25      A + G +G T Y+A + + G E E+ G IT WQ+ G ++ D +G+ +N
Sbjct: 565 AQSTGTPIPGSNGETAYKAVDGTVSKGVEFELNGAITDNWQLTFGATRYIAEDNEGNAVN 624
Query: 609 PDSVPERSEFKLFTAYHLAPEAPSGRTIGAGVRRQGETHTDPAALRIPNPAAKARAVANSR 668
30      P ++P + K+FT+Y L P P T+G GV Q +TD P RA
Sbjct: 625 P-NLPRTTVKMFETSYRL-PVMPE-LTVGGGVNWQNRVYTDTV-----TPYGTFR-----E 672
Query: 669 QKAYAVADIMARYRFNPRTELSLNVNDLNFKNHYRTQPDHR-SYGALRTVNAAFTYRF 724
35      Q +YA+ D+ RY+ L NV+NLE+K Y T + YG R + TY+F
Sbjct: 673 QGSYALVDLFTYQVTKNFSLQGNVNNLFDKTYDTNVEGSIVYGTNRNFSITGTYQF 729

```

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF23-1 (77.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 15A shows the results of affinity purification of the His-fusion protein, and Figure 15B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 15C) and for ELISA (positive result). These experiments confirm that ORF23-1 is a surface-exposed protein, and that it is a useful immunogen.

#### 40 Example 80

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 673>:

```

1  ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
51  GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTGCGCG GGAACGGCAA
101 TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTC
45 151  AGCGTCAGcA CGCCTGCTTC GGCGgcGgCa ATCATACCTT CGTCTTCGGA
201  AACGGGGATA AACGcGCCAC TCAAACCCCC GACCGCGCTG GAAGCCATCA
251  TGCCGCCTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTGTG
301  CCGTGCCTAC CGCAGACGCT CAAGCCCATT TnTTCAAGAA TCGGTGCCAC
351  TnAGTCGCCG ACGGGG..

```

50 This corresponds to the amino acid sequence <SEQ ID 674; ORF24>:

```

1  MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTE QTAVMASSLS
51  SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAVV
101  PCVPQTLKPI XSRMRATXSP TG..

```

Further work revealed the complete nucleotide sequence <SEQ ID 675>:

```

      1  ATGCGCACGG  CAGTGGTTTT  GCTGTTGATC  ATGCCGATGG  CGGCTTCGTC
     51  GGCAATGATG  CCGGAAATGG  TGTGCGCGGG  CGTGTGCGCG  GGAACGGCAA
    101  TCATATCCAA  GCCGACCGAA  CAAACGGCGG  TCATGGCTTC  GAGTTTGTCC
    151  AGCGTCAGCA  CGCCTGCTTC  GGCGGCGGCA  ATCATACCTT  CGTCTTCGGA
    201  AACGGGGATA  AACGCGCCAC  TCAAACCCCC  GACCGCGCTG  GAAGCCATCA
    251  TGCCGCCTTT  TTTCACGGCA  TCGTTCAGCA  ATGCCAAAGC  TGCTGTTGTG
    301  CCGTGCCTAC  CGCAGACGCT  CAAGCCCATT  TCTTCAAGAA  TGCGTGCCAC
    351  TGAGTCGCCG  ACGGCGGGGG  TCGGCGCCAG  CGACAAGTCG  AGAATACCAA
    401  ACGGGATATT  CAGCATTTTT  GAGGCTTCGC  GGCCGATGAG  TTCGCCACG
    451  CGGGTAATTT  TGAAAGCAGT  TTTCTTCACT  ACTTCCGCAA  CTTCGGTCAA
    501  TGTCGTTGCA  TCTGAATTTT  CCAACGCGGC  TTTTACGACA  CCTGGGCCGG
    551  ATACGCCGAC  ATTGATAACG  GCATCCGCTT  CGCCCGAACC  ATGAAACGCG
    601  CCCGCCATAA  ACGGGTTGTC  TTCCACCGCG  TTGCAGAAAC  CGACAATTTT
    651  AGCGCAGCCG  AAACCTTCGG  GCGTGATTTC  CGCCGTGCGT  TTGACGGTTT
    701  CGCCCGCCAG  CTTGACCGCA  TCCATATTGA  TACCGGCACG  CGTACTGCCG
    751  ATATTGATGG  AGCTGCACAC  AATATCGGTA  GTCTTCATCG  CTTCGGGAAT
    801  GGAGCGGATT  AACACCTCAT  CCGAAGGCGA  CATCCCTTTT  TGCACCAACG
    851  CGGAAAAACC  GCCGATAAAA  GACACACCGA  TGGCTTTGGC  AGCTTTATCC
    901  AAAGTTTGCG  CCACGCTGAC  GTAA

```

This corresponds to the amino acid sequence <SEQ ID 676; ORF24-1>:

```

      1  MRTAVVLLLI  MPMAASSAMM  PEMVCAGVSP  GTAIISKPTE  QTAVMASSLS
     51  SVSTPASAAA  IIPSSSETGI  NAPLKPPTAL  EAIMPPFFTA  SFSNAKAADV
    101  PCVPQTLKPI  SSRMRATESP  TAGVGASDKS  RIPNGIFSIF  EASRPMSPT
    151  RVILKAVFFT  TSATSVNVVA  SEFSNAAFTT  PGPDTPTLIT  ASASPEP*NA
    201  PAINGLSSTA  LQNTTILAQP  KPSGVISAVR  LTVSPASLTA  SILIPARVLP
    251  ILMELHTISV  VFASGMERI  NTSSEGDIPF  CTNAEKPPIK  DTPMALAALS
    301  KVCATLT*

```

Computer analysis of this amino acid sequence gave the following results:

### 30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF24 shows 96.4% identity over a 307 aa overlap with an ORF (ORF24a) from strain A of *N. meningitidis*:

```

      10      20      30      40      50      60
    orf24a.pep  MRTAVVLLLIMPMMAASSAMMPMVCAVSPGTAIISXPTEQTAVIASLSSNVSTPASAAA
    orf24       MRTAVVLLLIMPMMAASSAMMPMVCAVSPGTAIISKPTEQTAVMASSLSSVSTPASAAA
      10      20      30      40      50      60
      70      80      90     100     110     120
    orf24a.pep  IIPSSSXTGINAPLKPPTALEAIMPPFFTASFSNAKAADVPCVPQTLKPISSRMRAESP
    orf24       IIPSSSETGINAPLKPPTALEAIMPPFFTASFSNAKAADVPCVPQTLKPISSRMRAESP
      70      80      90     100     110     120
      130     140     150     160     170     180
    orf24a.pep  TAGVGASDKSRIPNGIFSIFEASRPMSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
    orf24       TAGVGASDKSRIPNGIFSIFEASRPMSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
      130     140     150     160     170     180
      190     200     210     220     230     240
    orf24a.pep  PGPDTPTLITASASPEPXNAPAIKGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA
    orf24       PGPDTPTLITASASPEPXNAPAIKGLSSXALQNTTILAQPKPSGVISAVRLTVSPASLTA
      190     200     210     220     230     240
      250     260     270     280     290     300
    orf24a.pep  SILIPARVLPILMELHTISVVFASGMERXNTSSEGDIPFCTSAEKPPIKDTPMALAALS
    orf24       SILIPARVLPILMELHTISVVFASGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS
      250     260     270     280     290     300

```

```

orf24a.pep  KVCATLTX
            |||||
orf24       KVCATLTX

```

5 The complete length ORF24a nucleotide sequence <SEQ ID 677> is:

```

1  ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
51 GGCAATGATG CCGGAAATGG TGTGCGCGGG TGTGTCGCCG GGAACGGCAA
101 TCATATCCAA NCCGACCGAA CAAACGGCGG TCATCGCTTC GAGTTTATCC
151 AACGTCAGCA CGCCTGCTTC GGCGGCGGCA ATCATACCTT CGTCTTCGGA
201 NACGGGGATA AACGCGCCAC TCAAACCGCC AACCGCGCTC GAAGCCATCA
251 TGCCGCCCTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
301 CCGTGCGTAC CGCAGACGCT CAAACCCATT TCTTCAAGAA TGC GCGCCAC
351 CGAGTCGCCG ACGGCAGGGG TCGGTGCCAG CGACAAGTCG AGAATACCAA
401 ACGGGATATT CAGCATTTTT GAGGCTTCGC GGCCGATGAG TTCGCCACG
15 451 CGGGTAATTT TGAAGGCGGT TTTCTTCACA ACTTCGGCAA CTTGGGTCAA
501 TGTCGTTGCA TCCGAATTTT CCAACGCGGC TTTTACGACA CCCGGGCCGG
551 ATACGCCGAC ATTAATCACA GCATCCGCTT CGCCTGAGCC GTGAAACGCG
601 CCCGCCATAN ACGGGTTGTC TTCCNCCGCG TTGCAGAACA CGACGATTTT
651 GGCGCAGCCG AAACCTTCTA GTGTGATTTC ANCCGTGCGT TTGATGGTTT
20 701 CGCCCGCCAG TCTGACCGCG TCCATATTGA TACCGCGCGC CGTACTGCCG
751 ATATTGATGG AGCTGCACAC GATATCAGTA GTCTTCATCG CTTCGGGAAT
801 GGAACGGATN AACACCTCGT CAGAAGGCGA CATACTTTT TGCACCAGCG
851 CGGAAAAGCC GCCAATAAAA GACACGCCGA TGGCTTTGGC AGCCTTATCC
901 AAAGTTTGCG CCACGCTGAC GTAA

```

25 This encodes a protein having amino acid sequence <SEQ ID 678>:

```

1  MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISXPTE QTAVIASSLS
51 NVSTPASAAA IIPSSSXTGI NAPLKPPTAL EAIMPPFFTA SFSNAKAADV
101 PCVPQTLKPI SSRMRATESP TAGVGASDKS RIPNGIFSIF EASRPMSSPT
151 RVILKAVFFT TSATSVNVVA SEFSNAAFTT PGPDPTLIT ASASPEP*NA
201 PAIXGLSSXA LQNTTILAQP KPSSVISXVR LMVSPASLTA SILIPARVLP
251 ILMELHTISV VFIASGMERX NTSSEGDIPF CTSAEKPPIK DTPMALAALS
301 KVCATLT*

```

It should be noted that this protein includes a stop codon at position 198.

ORF24a and ORF24-1 show 96.4% identity in 307 aa overlap:

```

35      10      20      30      40      50      60
orf24a.pep  MRTAVVLLLIMPMAASSAMMPMVCAGVSPGTAIISXPTEQTAVIASSLSNVSTPASAAA
orf24-1     MRTAVVLLLIMPMAASSAMMPMVCAGVSPGTAIISKPTQTAVMASSLSVSTPASAAA
            10      20      30      40      50      60
40      70      80      90      100     110     120
orf24a.pep  IIPSSSXTGINAPLKPPTALEAIMPPFFTASFNSNAKAADVPCVPQTLKPISSRMRATESP
orf24-1     IIPSSSETGINAPLKPPTALEAIMPPFFTASFNSNAKAADVPCVPQTLKPISSRMRATESP
            70      80      90      100     110     120
45      130     140     150     160     170     180
orf24a.pep  TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
orf24-1     TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
            130     140     150     160     170     180
50      190     200     210     220     230     240
orf24a.pep  PGPDPTLITASASPEPXNAPAIIXGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA
orf24-1     PGPDPTLITASASPEPXNAPAINGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA
            190     200     210     220     230     240
55      250     260     270     280     290     300
orf24a.pep  SILIPARVLPILMELHTISVVFIASGMERXNTSSEGDIPFCTSAEKPPIKDTPMALAALS
orf24-1     SILIPARVLPILMELHTISVVFIASGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS
            250     260     270     280     290     300
60

```

orf24a.pep KVCATLTX  
 5 orf24-1 KVCATLTX

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF24 shows 96.7% identity over a 121 aa overlap with a predicted ORF (ORF24ng) from *N.gonorrhoeae*:

10	orf24.pep	MRTAVVLLLIMPMAASSAMMPMVCAGVSPGTAIISKPTAQAVMASSLSVSTPASAAA	60
	orf24ng	MRTAVVLLLIMPMAASSAMMPMVCAGVSPGTAIMSKPTAQAVMASSLSVNTPASAAA	60
15	orf24.pep	IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVPCVPQTLKPIXSRMRATXSP	120
	orf24ng	IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVPCVPQTLKPISSRMRATESP	120
	orf24.pep	TG	122
20	orf24ng	TAGVGASDKSRMPNGIFSIFEASRPMSSPTRVILKAVFFTTTSATSVRLTASEFSSAALT	180

The complete length ORF24ng nucleotide sequence <SEQ ID 679> is:

1	ATGCGCACGG	CGGTGGTTTT	GCTGTTGATC	ATGCCGATGG	CGGCTTCGTC
51	GGCGATGATG	CCGGAATGG	TGTGCGCGGG	CGTGTGCGCG	GGAACGGCAA
101	TCATGTCCAA	ACCAACGGAG	CAGACGGCGG	TCATGGCTTC	GAGTTTGTCC
25	151	AGCGTCAACA	CGCCTGCCTC	GGCGGCGGCA	ATCATACTT
	201	AACGGGGATA	AACGCGCCGC	TCAAACCGCC	GACCGCGCTG
	251	TGCCGCCCTT	TTTCACGCA	TCGTTAGCA	ATGCCAAAGC
	301	CCGTGCGTAC	CGCAGACGCT	CAAGCCCAT	TCTTCAAGAA
	351	CGAGTCGCGG	ACGGCGGGGG	TCGGTGCCAG	CGACAAATCG
30	401	ACGGGATATT	CAGCATTTTT	GAGGCTTCG	GACCGATGAG
	451	CGGGTGATTT	TGAAAGCGGT	TTTCTTCAC	ACTTCGGCGA
	501	GCTGACCGCG	TCCGAATTTT	CCAGCGCGGC	TTTGACCACG
	551	ATACGCCGAC	ATTAATCACA	GCATCCGCTT	CGCCCGAGCC
	601	CCCGCCATAA	ACGGATTGTC	TTCCACCGCG	TTGCAGAACA
35	651	GGCGCAGCCG	AAACCTTCGG	GTGTGATTT	AGCCGTGCGT
	701	CGCCTGCCAG	CTTGACCGCA	TCCATATTGA	TACCGGCACG
	751	ATATTGATGG	AGCTGCACAC	GATATCGGTA	GTTTTCATCG
	801	GGAACGGATC	AACACCTCAT	CCGAAGGCGA	CATACCTTTT
	851	CGGAAAAGCC	GCCGATAAAG	GACACGCCGA	TGGCTTTGGC
40	901	AAAGTCTGCG	CCACGCTGAC	ATAA	

This encodes a protein having amino acid sequence <SEQ ID 680>:

1	MRTAVVLLLI	MPMAASSAMM	PEMVCAGVSP	GTAIMSKPTE	QTAVMASSLS
51	SVNTPASAAA	IIPSSSETGI	NAPLKPPTAL	EAIMPPFFTA	SFSNAKAAV
101	PCVPQTLKPI	SSRMRATESP	TAGVGASDKS	RMPNGIFSIF	EASRPMSSPT
45	151	RVILKAVFFT	TSATSVRLTA	SEFSSAALT	PGPDPTLIT
	201	PAINGLSSTA	LQNTTILAQP	KPSGVISAVR	LMVSPASLTA
	251	ILMELHTISV	VFIASGTERI	NTSSEGDIPF	CTSAEKPPIK
	301	KVCATLT*			DTPMALAALS

ORF24ng and ORF24-1 show 96.1% identity in 307 aa overlap:

50	orf24-1.pep	MRTAVVLLLIMPMAASSAMMPMVCAGVSPGTAIISKPTAQAVMASSLSVSTPASAAA	60
	orf24ng	MRTAVVLLLIMPMAASSAMMPMVCAGVSPGTAIMSKPTAQAVMASSLSVNTPASAAA	60
55	orf24-1.pep	IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVPCVPQTLKPISSRMRATESP	120
	orf24ng	IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVPCVPQTLKPISSRMRATESP	120

-390-

		130	140	150	160	170	180
	orf24-1.pep	TAGVGASDKSRIPNGIFSIFEASRPMSPTRVILKAVFFTTTSATSVNVVASEFSNAFFT					
	orf24ng	TAGVGASDKSRMPNGIFSIFEASRPMSPTRVILKAVFFTTTSATSVRLTASEFSSAALT					
5		130	140	150	160	170	180
	orf24-1.pep	PGPDTPTLITASASPEPKXNAPAINGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA					
10	orf24ng	PGPDTPTLITASASPEPWNAPAINGLSSTALQNTTILAQPKPSGVISAVRLMVSPASLTA					
		190	200	210	220	230	240
	orf24-1.pep	SILIPARVLPILMELHTISVVFASGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS					
15	orf24ng	SILIPARVLPILMELHTISVVFASGTERINTSSEGDIPFCTSAEKPPIKDTPMALAALS					
		250	260	270	280	290	300
20	orf24-1.pep	KVCATLT					
	orf24ng	KVCATLT					

Based on this analysis, including the presence of a putative leader sequence (first 18 aa – double-underlined) and putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 81

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 681>:

```

30      1  ..ACCGACGTGC AAAAAGAGTT GGTCCGCGAA CAACGCAAGT GGGCGCAGGA
      51  AAAAATCAGC AACTGCCGAC AAGCCGCCGC GCAGGCAGAC CGGCAGGAAT
     101  ACGCCGAATA CCTCAAGCTG CAATCGCACA CGCGGATGAC GCGCGAACGG
     151  ATACAGTATC TTCGCGGCTA TTCCATCGAT TAG

```

This corresponds to the amino acid sequence <SEQ ID 682; ORF25>:

```

35      1  ..TDVQKELVGE QRKWAQEKIS NCRQAAQAD RQEYAEYLKL QCDTRMTRER
     51  IQYLRGYSID *

```

Further work revealed the complete nucleotide sequence <SEQ ID 683>:

```

40      1  ATGTATCGGA AACTCATTGC GCTGCCGTTT GCCCTGCTGC TTGCCGCTTG
     51  CGGCAGGGAA GAACCGCCCA AGGCATTGGA ATGCGCCAAC CCCGCCGTGT
     101  TGCAAGGCAT ACGCGGCAAT ATTCAAGAAA CGCTCACGCA GGAAGCGCGT
     151  TCTTTCGCGC GCGAAGACGG CAGGCAGTTT GTCGATGCCG ACAAATTAT
     201  CGCCGCCGCC TACGGTTTGG CGTTTTCTTT GGAACACGCT TCGGAAACGC
     251  AGGAAGGCGG GCGCACGTTT TGTATCGCCG ATTTGAACAT TACCGTGCCG
     301  TCTGAAACGC TTGCCGATGC CAAGGCAAAC AGCCCCCTGT TGTACGGGGA
45     351  AACTGCTTTG TCGGATATTG TGCGGCAGAA GACGGGCGGC AATGTGAGT
     401  TTAAGACGG CGTATTGACG GCAGCGTCC GCTTCCTGCC CGTCAAAGAC
     451  GGTACAGACG CATTTGTCTGA CAACACGGTC GGTATGGCGG CGCAAACGCT
     501  GTCTGCCGCG CTGCTGCCTT ACGGCGTGAA GAGCATCGTG ATGATAGACG
     551  GCAAGGCGGT GAAAAAGAA GACGCGGTCA GGATTTGAG CGGAAAAGCC
50     601  CGTGAAGAAG AACCGTCCAA ACCACGCCC GAAGACATT TGAACACAA
     651  TGCCGCCGGC GCGGATGCGG GCGTACCCA AGCCGCAGAA GCGCGCCCG
     701  AACCGGAAAT CCTGCATCCT GACGACGGCG AGCGTGCCGA TACCGTTACC
     751  GTATCACGGG GCGAAGTGGA AGAGGCGCGC GTACAAAACC AGCGTGCGGA
     801  ATCCGAAATT ACCAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAG
55     851  AGTTGGTCGG CGAACAACGC AAGTGGGCGC AGGAAAAAAT CAGCAACTGC
     901  CGACAAGCCG CCGCGCAGGC AGACCGGCAG GAATACGCCG AATACCTCAA
     951  GCTGCAATGC GACACGCGGA TGACGCGCGA ACGGATACAG TATCTTCGCG
    1001  GCTATTCCAT CGATTAG

```

This corresponds to the amino acid sequence <SEQ ID 684; ORF25-1>:

```

      1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQGIRGN IQETLTQEAR
     51 SFAREDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP
    101 SETLADAKAN SPLLYGETAL SDIVRQKTGG NVEFKDGVLT AAVRFLPVKD
    151 GQTAFVDNTV GMAAQTL SAA LLPGVK SIV MIDGKAVKKE DAVRILSGKA
    201 REEPPSKPTP EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDGERADTVT
    251 VSRGEVEEAR VQNQRAESEI TKLWGGLD TD VQKELVGEQR KWAQEKISNC
    301 RQAAAQADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID*
  
```

Computer analysis of this amino acid sequence gave the following results:

#### 10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF25 shows 98.3% identity over a 60aa overlap with an ORF (ORF25a) from strain A of *N.*

*meningitidis*:

```

      10      20      30
    orf25.pep      TDVQKELVGEQRKWAQEKISNCRQAAAQAD
    15      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    orf25a      VTVSRGEVEEARVQNQRAESEITKLWGGLD TDVQKELVGEXRKWAQEKISNCRQAAAQAD
      250      260      270      280      290      300

      40      50      60
    20      orf25.pep      RQEYAEYLKLCQDTRMTRERIQYLRGYSIDX
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    orf25a      RQEYAEYLKLCQDTRMTRERIQYLRGYSIDX
      310      320      330
  
```

The complete length ORF25a nucleotide sequence <SEQ ID 685> is:

```

    25      1 ATGTATCGGA AACTCATTGC GCTGCCGTTT GCCCTGCTGC TTGCCGCTTG
      51 CGGCAGGGAA GAACCGCCCA AGGCATTGGA ATGCGCCAAC CCCGCCGTGT
    101 TGCAANGCAT ACGCNGCAAT ATTCAGGAAA CGCTCACGCA GGAAGCGCGT
    151 TCTTTCGCGC GCGAAGACNG CANGCAGTTT GTCGATGCCG ACNAAATTAT
    201 CGCCGCCGCC TANGNTNNGN NGNTNTCTTT GGAACACGCT TCGGAAACGC
    301 AGGAAGGCGG GCGCACGTTT TGTNTCGCCG ATTTGAACAT TACCGTGCCG
    351 TCTGAAACGC TTGCCGATGC CAAGGCAAAC AGCCCCCTGC TGTACGGGGA
    401 AACCGCTTTG TCGGATATTG TGCGGCAGAA GACGGGCGGC AATGTCGAGT
    451 TTAAGACGGC CGTATTGACG GCAGCCGTCC GCTTCTTACC CGTCAAAGAC
    501 GGTGAGANGG CATTTGTGCA CAACACGGTC GGTATGGCGG CGCAAACGCT
    551 GTCTGCCGCG TTGCTGCCTT ACGGCGTGAA GAGCATCGTG ATGATAGACG
    601 GCAAGGCGGT AAAAAAGAA GACGCGGTCA GGATTNTGAG CNGANAAGCC
    651 CGTGAANAAG AACCGTCCAA ANCCNNGCCC GAAGACATTT TGGAACATAA
    701 TGCCGCCGGA GGGGATGCAG ACGTACCCA AGCCGGAGAA GACGCGCCCG
    751 AACCGGAAAT CCTGCATCCT GACGACGGCG AGCGTGCCGA TACCGTTACC
    801 GTATCACGGG GCGAAGTGGA AGAGGCGCGN GTACAAAACC AGCGTGCGGA
    851 ATCCGAAATT ACCAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAG
    901 AGTTGGTCGG CGAANAACGC AAGTGGGCGC AGGAAAAAAT CAGCAACTGC
    951 CGACAAGCCG CCGCGCAGGC AGACGGGCAG GAATACGCCG AATACCTCAA
    1001 GCTGCAATGC GACACGCGGA TGACGCGCGA ACGGATACAG TATCTTCGCG
  
```

This encodes a protein having amino acid sequence <SEQ ID 686>:

```

      1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQXIRXN IQETLTQEAR
     51 SFAREDXXQF VDADKIIAAA XXXXSLEHA SETQEGGRTF CXADLNITVP
    101 SETLADAKAN SPLLYGETAL SDIVRQKTGG NVEFKDGVLT AAVRFLPVKD
    151 GQXAFVDNTV GMAAQTL SAA LLPGVK SIV MIDGKAVKKE DAVRIXSXXA
    201 REXEPSKXXP EDILEHNAAG GDADV PQAGE DAPEPEILHP DDGERADTVT
    251 VSRGEVEEAR VQNQRAESEI TKLWGGLD TD VQKELVGEQR KWAQEKISNC
    301 RQAAAQADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID*
  
```

ORF25a and ORF25-1 show 93.5% identity in 338 aa overlap:

```

    55      10      20      30      40      50      60
    orf25a.pep      MYRKLIALPFALLLAACGRE EPPKALECAN PAVLQXIRXNIQETLTQEARSFAREDXXQF
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    orf25-1      MYRKLIALPFALLLAACGRE EPPKALECAN PAVLQGIRGNIQETLTQEARSFAREDGRQF
  
```



-392-

		10	20	30	40	50	60
5	orf25a.pep	70	80	90	100	110	120
	orf25-1	70	80	90	100	110	120
10	orf25a.pep	130	140	150	160	170	180
	orf25-1	130	140	150	160	170	180
15	orf25a.pep	190	200	210	220	230	240
	orf25-1	190	200	210	220	230	240
20	orf25a.pep	250	260	270	280	290	300
	orf25-1	250	260	270	280	290	300
25	orf25a.pep	310	320	330	339		
	orf25-1	310	320	330			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF25 shows 100% identity over a 60aa overlap with a predicted ORF (ORF25ng) from

35 *N.gonorrhoeae*:

orf25.pep	TDVQKELVGEQRKWAQEKISNCRQAAAQAD	30
orf25ng	VTVSRGEVEEARVQNQRAESEITKLWGGLD TDVQKELVGEQRKWAQEKISNCRQAAAQAD	308
orf25.pep	RQEYAEYKLQCDTRMTRERIQLRGYSID	60
orf25ng	RQEYAEYKLQCDTRMTRERIQLRGYSID	338

The complete length ORF25ng nucleotide sequence <SEQ ID 687> is:

45	1	ATGTATCGGA	AACTCATTGC	GCTGCCGTTT	GCCCTGCTGC	TTGCAGCGTG
	51	CGGCAGGGAA	GAACCGCCCA	AGGCGTTGGA	ATGCGCCAAC	CCCGCCGTGT
	101	TGCAGGACAT	ACGCGGCAGT	ATTCAGGAAA	CGCTCACGCA	GGAAGCGCGT
	151	TCTTTCGCGC	GCGAAGACGG	CAGGCAGTTT	GTCGATGCCG	ACAAAATTAT
	201	CGCCGCCGCC	TACGGTTTGG	CGTTTCTTT	GGAACACGCT	TCGGAAACGC
50	251	AGGAAGGCGG	GCGCACGTTT	TGTATCGCCG	ATTTGAACAT	TACCGTGCCG
	301	TCTGAAACGC	TTGCCGATGC	CGAGGCAAAC	AGCCCCCTGC	TGTATGGGGA
	351	AACGTCTTTG	GCAGACATCG	TGCAGCAGAA	GACGGGCGGC	AATGTCGAGT
	401	TTAAAGACGG	CGTATTGACG	GCAGCCGTCC	GCTTCCTGCC	CGCCAAAGAC
	451	GCTCGGACGG	CATTTATCGA	CAACACGGTC	GGTATGGCGA	CGCAAACGCT
55	501	GTCTGCCGCG	TTGCTGCCCT	ACGGCGTGAA	GAGCATCGTG	ATGATAGACG
	551	GCAAGGCGGT	GACAAAAGAA	GACGCGGTCA	GGGTTTGTAG	CGGCAAAGCC
	601	CGTGAAGAAG	AACCGTCCAA	ACCCACCCCG	GAAGACATTT	TGGAACACAA
	651	TGCCGCCGGC	GGCGATGCGG	GCGTACCCCA	AGCCGCAGAA	GGCGCACCCG
	701	AACCCGAAAT	CCTGCATCCC	GACGACGTCG	AGCGTGCCGA	TACCGTTACC
60	751	GTATCACGGG	GCGAAGTGGA	AGAGGCGCGC	GTACAAAACC	AACGTGCGGA
	801	ATCCGAAATT	ACCAAACTTT	GGGGAGGACT	CGATACCGAC	GTGCAAAAAG
	851	AGTTGGTCGG	CGAACAGCGC	AAGTGGGCGC	AGGAAAAAAT	CAGcaactgc
	901	cgACAAGCCG	CCGCGCAGGC	AGACGGGCAG	GAATACGCCG	AATACCTCAA
	951	GCTCCAATGC	GACACGCGGA	TGACGCGCGA	ACggaTACAG	TATCTTCGCG
	1001	GCTATTCCAT	CGATTAG			

65 This encodes a protein having amino acid sequence <SEQ ID 688>:

-393-

1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQDIRGS IQETLTQEAR  
 51 SFAREDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP  
 101 SETLADAEAN SPLLYGETSL ADIVQKTGG NVEFKDGVLT AAVRFLPAKD  
 151 ARTAFIDNTV GMATQTLCAA LPPYGVKSIV MIDGKAVTKE DAVRVLGSKA  
 5 REEPPSKPTP EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDVERADTVT  
 201 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEQR KWAQEKISNC  
 251 RQAAAQADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID\*  
 301

ORF25ng and ORF25-1 show 95.9% identity in 338 aa overlap:

10	orf25-1.pep	10	20	30	40	50	60
	orf25ng	10	20	30	40	50	60
15	orf25-1.pep	70	80	90	100	110	120
	orf25ng	70	80	90	100	110	120
20	orf25-1.pep	130	140	150	160	170	180
	orf25ng	130	140	150	160	170	180
25	orf25-1.pep	190	200	210	220	230	240
	orf25ng	190	200	210	220	230	240
30	orf25-1.pep	250	260	270	280	290	300
	orf25ng	250	260	270	280	290	300
35	orf25-1.pep	310	320	330	339		
	orf25ng	310	320	330			

45 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

50 ORF25-1 (37kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 16A shows the results of affinity purification of the GST-fusion protein, and Figure 16B shows the results of expression of the His-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 16C), ELISA (positive result), and FACS analysis (Figure 16D). These experiments confirm that ORF25-1 is a surface-exposed protein, and  
 55 that it is a useful immunogen.

Figure 16E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF25-1.

### Example 82

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 689>

```

5      1  ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
      51  TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
     101  GCATCGGTAT TCTGGwysGC GTTGCCCTTTT TGGTCGGCGG CAACCCCGTC
     151  GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
     201  CGsyGATTGG TCGCTGGGCA AACC AAAAAT CTTGGTTTTT CkGATACTTT
     251  TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA T.....
10      //
      851  .....AC TTCGCTGGTA
     901  TTCGGCGGCA CTTGCGGCGT CTTTGCCGTC GTTCTCTGCA CGCTCGGCAC
     951  GATTAAAACC GCCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
    1001  TGTTCCGGCGC AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
15    1051  GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA
     1101  CATCCATCCC GGCTTCCCTG CCGTCATCCT CTTCTGCTC GCCAGCGTGA
     1151  TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
     1201  ATTGCCGCCG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
     1251  TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA
20    1301  TTTCCGACAC GACCATCTG TCGTCCACCG GCGCGCGCTG CAACCACATC
     1351  GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCGCGCGC
     1401  CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
     1451  TTGGCAGCAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT
    1501  AAAAAA..

```

25 This corresponds to the amino acid sequence <SEQ ID 690; ORF26>:

```

      1  MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILXX VAFVLVGGNPV
     51  DGLTHLKDMV VGLAWSDXDW SLGKPKILVF XILLGIFTSL LTYSGSN...
      //
    251  .....TSLV
    301  FGGTCGVFAV VLCTLGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV
    351  VGEMHTGDYL STLVAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP
    401  IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDITIL SSTGARNHI
    451  DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTGIV LAVLIFLLKD
    501  KK..

```

35 Further work revealed the complete nucleotide sequence <SEQ ID 691>:

```

      1  ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
     51  TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
    101  GCATCGGTAT TCTGGTCGCG GTTGCCCTTTT TGGTCGGCGG CAACCCCGTC
    151  GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
40    201  CGGCGATTGG TCGCTGGGCA AACC AAAAAT CTTGGTTTTT CTGATACTTT
     251  TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT
     301  GCCGACTGGG CAAAACGGCA CATTAAAAAC CGGCGCGGCG CGAAAATGCT
     351  GACCGCCTGC CTCGTGTTTC TAACCTTTAT CGACGACTAT TTCACAGTC
     401  TCGCCGTCGG TGCGATTGCC CGCCCCGTTA CCGACAAGTT TAAAGTTTCC
45    451  CGCACCAAAC TCGCTACAT CCTCGACTCC ACTGCCGCTC CTATGTGCGT
     501  GCTGATGCCC GTTTCAAGCT GGGGCGCGTC GATTATCGCC ACGCTTGCCG
     551  GACTGCTCGT TACCTACAAA ATCACCGAAT ACACGCCGAT GGGGACGTTT
     601  GTCGCCATGA GCCTGATGAA CTATTACGCA CTGTTTGCCC TGATTATGGT
50    651  GTTCGTCGTC GCATGGTTTT CTTTCGACAT CGGCTCGATG GCACGTTTTCG
     701  AACAAAGCCG GTTGAACGAA GCCCACGATG AACTGCGCTT TTCAGACGCT
     751  ACCAAAGGTC GTGTTTACGC ACTGATTATT CCCGTTTTGG CCTTAATCGC
     801  CTCAACGGTT TCCGCCATGA TCTACACCGG CGCGCAGGCA AGCGAAACCT
     851  TCAGCATTTT GGGGGCATT GAAAACACG ACGTAAACAC TTCGTGTGTA
55    901  TTCGGCGGCA CTTGCGGCGT CTTTGCCGTC GTTCTCTGCA CGCTCGGCAC
     951  GATTAAAACC GCCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
    1001  TGTTCCGGCGC AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
    1051  GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA
    1101  CATCCATCCC GGCTTCCCTG CCGTCATCCT CTTCTGCTC GCCAGCGTGA
    1151  TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
60    1201  ATTGCCGCCG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
     1251  TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA
    1301  TTTCCGACAC GACCATCTG TCGTCCACCG GCGCGCGCTG CAACCACATC

```

1351 GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC  
 1401 CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT  
 1451 TTGGCAGGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT  
 1501 AAAAAACGCG CCAACGCCTG A

5 This corresponds to the amino acid sequence <SEQ ID 692; ORF26-1>:

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFLVGGNPV  
 51 DGLTHLKDMV VGLAWSGDGW SLGKPKILVF LILLGIFTSL LTYSGSNQAF  
 101 ADWAKRHIKN RRGAKMLTAC LVFVTFIDY FHSLAVGAIA RPVTDKFKVS  
 151 RTKLAYILDS TAAPMCVLMV VSSWGASIIA TLAGLLVYK ITEYTPMGTF  
 10 201 VAMSLMNYA LFALIMVFVV AWFSFDIGSM ARFEQAALNE AHDETAVSDA  
 251 TKGRVYALII PVLALIASTV SAMIYTGAA SETFSILGAF ENTDVNTSLV  
 301 FGGTCGVLA VCLTGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV  
 351 VGEHMTGDYL STLVAAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP  
 401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDITIL SSTGARCNIH  
 15 451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD  
 501 KKRANA\*

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical transmembrane protein HI1586 of *H. influenzae* (accession number P44263)

ORF26 and HI1586 show 53% and 49% amino acid identity in 97 and 221 aa overlap at the

20 N-terminus and C-terminus, respectively:

Orf26 1 MQLIDYSHSFFSVVPPFLALALAVITRRVXXXXXXXXXXVAFLVGGNPVDGLTHLKDMV 60  
 M+LID+S S +S+VP LA+ LA+ TRRV L +L V  
 HI1586 14 MELIDFSSSVWSIVPALLAIILAIATRRVLVLSAGIIIGSLMLSDWQIGSAFNLYLVKNV 73

25 Orf26 61 VGLAWSDXDWSLGKPKILVFXILLGIFTSLTYSGSN 97  
 V L ++D + + I++F +LLG+ T+LLT SGSN  
 HI1586 74 VSLVYADGEIN-SNMNIVLFLLLLGLTALLTVSGSN 109

//

30 Orf26 86 IFTSLLTYSGS--NTSLVFGGTCGVFAVVLCTL--GTIKTADYPKAVWQGA KSMFGXXXX 141  
 +F+ L T+ + TSLV GG C + L + + +Y ++ G KSM G  
 HI1586 299 VFSVLGTFENTVVGTSLVVGGFCSIIISTLLIILDRQVSVPEYVRSWIVGIKSMGSAIAI 358

35 Orf26 142 XXXXXXSTVVGEHMTGDYLSTLVAGNIHPGFLPVILFLLASVMAFATGT SWGTFGIMLP 201  
 + +VG+M TG YLS+LV+GNI FLPVILF+L + MAF+TGTSWGTFGIMLP  
 HI1586 359 LFFAWTINKIVGDMQTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGT SWGTFGIMLP 418

40 Orf26 202 IAAAMAVKVEPALIIPCMSAVMAGAVCGDH CSPISDITILSSTGARCNIHDHVTSQXXXX 261  
 IAAAMA P L++PC+SAVMAGAVCGDH CSP+SDTITLSSTGA+CNHIDHVT+Q  
 HI1586 419 IAAAMAANAPELLLPCLSAVMAGAVCGDH CSPVSDTITLSSTGAKCNHIDHVTTQLPYA 478

Orf26 262 XXXXXXXXXXXXXXXXXXXKSALLGFGTTGIVLAVLIFLLKDK 302  
 S L GF T + L V+IF +K +

45 HI1586 479 ATVATATSIGYIVVGFTYSGLAGFAATAVSLIVIIFAVKKR 519

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF26 shows 58.2% identity over a 502aa overlap with an ORF (ORF26a) from strain A of *N. meningitidis*:

50 orf26.pep 10 20 30 40 50 60  
 MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXXVAFLVGGNPVDGLTHLKDMV  
 orf26a MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV

55 orf26.pep 70 80 90 99  
 VGLAWSDXDWSLGKPKILVFXILLGIFTSLTYSGSNXX-----  
 orf26a VGLAWSGDWDLGKPKILVFLILLGIFTSLTYSGSNQAFADWAKRHIKNRRGAKMLTAC

60 70 80 90 100 110 120

-396-

orf26.pep		-----
5	orf26a	<u>LVFVTFIDDYFHS</u> LAVGAXARPVTDKFKVSRAKLAYILDSTAAPMCVLMFVSSWGASIIA 130 140 150 160 170 180
orf26.pep		-----
10	orf26a	<u>TLAGLLV</u> TYKITEYTPMGTFFVAMSLMNYIALFALIMVFVVAWFSDIGSMARFEQAALNE 190 200 210 220 230 240
orf26.pep		-----100 110-----TSLV 
15	orf26a	AHDETAVSDGSGWRVYALII <u>PVLALIA</u> STVSAMIYTGAQASETFSILGAFENTDVNTSLV 250 260 270 280 290 300
orf26.pep		120 130 140 150 160 170 <u>EGGTCGVFAVVLCT</u> LGTIKTADYPKAVWQAKSMFGAIAILILAWLISTVVGEMHTGDYL      :
20	orf26a	<u>EGGTCGVLA</u> VVLCTLGTIKIADYPKAVWQAKSMFGAIAILILAWLISTVVGEMHTGDYL 310 320 330 340 350 360
orf26.pep		180 190 200 210 220 230 STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP <u>IAAAMAVKVEPALII</u> PCMSA      :
25	orf26a	STLVAGNIHPGFLXVILFLLASVMAFATGTSWGTFGIMLP <u>IAAAMAVKVDPSLI</u> IPCMSA 370 380 390 400 410 420
orf26.pep		240 250 260 270 280 290 <u>VMAGAVCGDH</u> CSPISDTTILSSTGARNHIDHVTSQLPYALTVA <u>AAAAASGYLALGLTK</u> SA 
30	orf26a	<u>VMAGAVCGDH</u> CSPISDTTILSSTGARNHIDHVTSQLPYALTVA <u>AAAAASGYLALGLTK</u> SA 430 440 450 460 470 480
orf26.pep		300 310 <u>LLGFGTTGIVL</u> AVLIFLLKDKK      :
35	orf26a	<u>LLGFGXTGIVL</u> AVLIFLLKDKKRANAX 490 500

The complete length ORF26a nucleotide sequence <SEQ ID 693> is:

45	1	ATGCAGCTGA	TCGACTATTC	ACATTCATTT	TTCTCGGTTG	TGCCACCCCTT
	51	TTTGGCACTG	GCACTTGCCG	TCATTACCCG	CCGCGTACTG	CTGTCTTTAG
	101	GCATCGGTAT	TCTGGTCGGC	GTTGCCTTTT	TGGTCGGCGG	CAACCCCGTC
	151	GACGGTCTGA	CACACCTGAA	AGACATGGTC	GTCGGCTTGG	CTTGGTCAGA
	201	CGGCGATTGG	TCGCTGGGCA	AACCAAAANT	CTTGGTTTTC	CTGATACTTT
	251	TGGGTATTTT	TACTTCCCTG	CTGACCTACT	CCGGCAGCAA	TCAGGCGTTT
50	301	GCCGACTGGG	CAAACGGCA	CATTAAAAAC	CGGCGGGCGG	CGAAATGCT
	351	GACCGCTGCG	CTCGTGTTCC	TAACCTTTAT	CGACGACTAT	TTCCACAGTC
	401	TCGCCGTCGG	TGCGNTTGCC	CGCCCCGTTA	CCGACAAGTT	TAAAGTTTCC
	451	CGCGCCAAAC	TCGCCTACAT	CCTCGACTCC	ACTGCCGCGC	CTATGTGCGT
	501	GCTGATGCCC	GTTTCAAGCT	GGGGCGCGTC	GATTATCGCC	ACGCTTGCCG
55	551	GACTGCTCGT	TACCTACAAA	ATCACCGAAT	ACACGCCGAT	GGGGACGTTT
	601	GTCGCCATGA	GCCTGATGAA	CTATTACGCA	CTGTTTGCCC	TGATTATGGT
	651	GTTCTGTCGT	GCATGGTTCT	CCTTCGACAT	CGGCTCGATG	GCACGTTTCG
	701	AACAAGCCGC	GTTGAACGAA	GCCCACGATG	AAACTGCCGT	TTCAGACGGC
	751	AGCTGGGGCA	GGGTTTACGC	ATTGATTATT	CCCGTTTTGG	CCTTAATCGC
60	801	CTCAACGGTT	TCCGCCATGA	TCTACACCGG	TGCACAGGCA	AGCGAAACCT
	851	TCAGCATTTT	GGGTGCATTT	GAAAATACGG	ACGTGAACAC	TTCGCTGGTA
	901	TTCGGCGGCA	CTTGCGGCGT	GCTTGCCGTC	GTCTCTGCA	CGCTCGGCAC
	951	GATTAAAAATC	GCCGATTATC	CCAAAGCCGT	TTGGCAGGGT	GCGAAATCCA
	1001	TGTTCCGGCGC	AATCGCCATT	TTAATCCTTG	CCTGGCTCAT	CAGTACGGTT
65	1051	GTCGGCGAAA	TGCACACAGG	CGACTACCTC	TCCACGCTGG	TTGCGGGCAA
	1101	CATCCATCCC	GGCTTCCTGN	CCGTATCCTT	TTTCCTGCTC	GCCAGCGTGA
	1151	TGGCGTTTGC	CACAGGCACA	AGCTGGGGGA	CGTTCGGCAT	CATGCTGCCG
	1201	ATTGCCGCCG	CCATGGCGGT	CAAAGTCGAT	CCCTCACTGA	TTATCCCGTG
	1251	TATGTCCGCC	GTGATGGCGG	GGGCGGTATG	CGGCGACCAC	TGCTCGCCCA
70	1301	TTTCCGACAC	GACCATCCTG	TCGTCCACCG	GCGCGCGCTG	CAACCACATC

-397-

1351 GACCACGTTA CNTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC  
 1401 CGCATCGGGN TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGTT  
 1451 TTGGCANGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT  
 1501 AAAAAACGCG CCAACGCCTG A

5 This encodes a protein having amino acid sequence <SEQ ID 694>:

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFVLGGNPV  
 51 DGLTHLKDMV VGLAWSGDW SLGKPKXLVF LILLGIFTSL LTYSGSNQAF  
 101 ADWAKRHIKN RRGAKMLTAC LVFVTFIDY FHSLAVGAXA RPVTDKFKVS  
 151 RAKLAYILDS TAAPMCVLMV VSSWGASIIA TLAGLLVITYK ITEYTPMGTF  
 10 201 VAMSLMNYIA LFALIMVFVW AWFSFDIGSM ARFEQAALNE AHDETAVSDG  
 251 SWGRVYALII PVLALIASTV SAMIYTGAQA SETFSILGAF ENTDVNTSLV  
 301 FGGTCGVLAV VLCTLTGTIKI ADYPKAVWQG AKSMFGAIAI LILAWLISTV  
 351 VGMHTGDYL STLAVAGNIHP GFLXVILFLL ASVMAFATGT SWGTFGIMLP  
 401 IAAAMAVKVD PSLIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARNHI  
 15 451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGXTGIV LAVLIFLLKD  
 501 KKRANA\*

ORF26a and ORF26-1 show 97.8% identity in 506 aa overlap:

		10	20	30	40	50	60
20	orf26a.pep	MQLIDYSHSF	FSVVPPFLAL	ALAVITRRVL	LSLGIGILVG	VAFVLGGNPV	DGLTHLKDMV
	orf26-1	MQLIDYSHSF	FSVVPPFLAL	ALAVITRRVL	LSLGIGILVG	VAFVLGGNPV	DGLTHLKDMV
		10	20	30	40	50	60
25	orf26a.pep	70	80	90	100	110	120
	orf26-1	VGLAWSGDW	SLGKPKXLV	FLILLGIFT	SLLTYSGSN	QAFADWAKR	HIKNRRGAK
		70	80	90	100	110	120
30	orf26a.pep	130	140	150	160	170	180
	orf26-1	LVFVTFIDY	FHSLAVGAX	ARPVTDKFK	VSRAKLAYI	L DSTAAPMC	VLMPVSSWG
		130	140	150	160	170	180
35	orf26a.pep	190	200	210	220	230	240
	orf26-1	TLAGLLVITY	KITEYTPMG	TFVAMSLMN	YALFALIMV	FVVAWFSFD	IGSMARFEQ
		190	200	210	220	230	240
40	orf26a.pep	250	260	270	280	290	300
	orf26-1	AHDETAVSD	G SWGRVYAL	IIPVLALIAS	TVSAMIYTGA	QASETF SIL	GAFENTDVNT
		250	260	270	280	290	300
45	orf26a.pep	310	320	330	340	350	360
	orf26-1	FGGTCGVLAV	VLCTLTGTI	KIADYPKAV	WQGA KSMFG	AIAILILAW	L ISTVVGMHT
		310	320	330	340	350	360
50	orf26a.pep	370	380	390	400	410	420
	orf26-1	STLVAGNIHP	GFLXVILFLL	ASVMAFATGT	SWGTFGIMLP	IAAAMAVKV	DP SLIIPCMS
		370	380	390	400	410	420
55	orf26a.pep	430	440	450	460	470	480
	orf26-1	VMAGAVCGDH	CSPISDTTIL	SSTGARNHI	DHVTSQLPYA	LTVAAAAASG	YLALGLTKSA
		430	440	450	460	470	480
60	orf26a.pep	490	500				
	orf26-1	LLGFGXTGIV	LAVLIFLLK	DKKRANAX			
		490	500				

-398-

```

      |||||:|||||
orf26-1  LLGFGTTGIVLAVLI FLLKDKKRANAX
              490       500

```

### 5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF26 shows 94.8% and 99% identity in 97 and 206 aa overlap at the N-terminus and C-terminus, respectively, with a predicted ORF (ORF26ng) from *N. gonorrhoeae*:

```

10 orf26.pep  MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXXVAFLVGGNPVDGLTHLKDMV 60
    orf26ng   MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV 60

    orf26.pep  VGLAWSDXDWSLKGPKILVFXILLGIFTSLLTYSGSN 97
    orf26ng   VGLAWADGDWSLKGPKILVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRCGAKMLTAC 120

15 //

    orf26.pep  TSLVFGGTCGVFAVVLCTLGTIKTADYPKA 326
    orf26ng   ASTVSAMIYTGAQASETFSILGAFENTDVNTSLVFGGTCGVLA VVLCFTGTIKTADYPKA 326

    orf26.pep  VWQGA KSMFGAIAILILAWLISTVVGEMHTGDYLS TLVAGNIHPGFLPVILFLLASVMAF 386
    orf26ng   VWQGA KSMFGAIAILILAWLISTVVGEMHTGDYLS TLVAGNIHPGFLPVILFLLASVMAF 386

25 orf26.pep  ATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPI SDTTILSSTGAR 446
    orf26ng   ATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPI SDTTILSSTGAR 446

30 orf26.pep  CNHIDHVTSQ LPYALTVA AAAASGYLALGLTKSALLGFGTTGIVLAVLI FLLKDKK 502
    orf26ng   CNHIDHVTSQ LPYALTVA AAAASGYLALGLTKSALLGFGTTGIVLAVLI FLLKDKKRADV 506

```

The complete length ORF26ng nucleotide sequence <SEQ ID 695> is:

```

35 1 ATGCAGCTGA TTGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
   51 TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
  101 GCATCGGTAT TTTGGTCGGC GTTGCCCTTTT TGGTCGCGCG CAACCCCGTC
  151 GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTGGGGCAGA
  201 CGGCGATTGG TCGCTGGGCA AACCAAAAAT CTTGGTTTTC CTGATACTTT
  251 TGGGCATTTT CACTTCACTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT
40 301 GCCGACTGGG CAAAACGGCA CATTAAAAAC CGGTGCGGCG CGAAAATGCT
   351 GACCGCTGCG CTGCTGTTTCG TAACCTTTAT CGACGACTAT TTCCACAGCC
   401 TCGCCGTCGG TCGGATTGCC CGCCCCGTTA CCGACAAGTT TAAAGTTTCC
   451 CGCGCCAAAC TCGCTACAT CCTCGACTCC ACTGCCTCGC CCATGTGCGT
   501 GCTGATGCCC GTTCAAGCT GGGGCGCGTC GATTATCGCC ACGCTTGCCG
45 551 GATTGCTCGT TACCTACAAA ATTACCGAAT ACACGCCGAT GGGGACGTTT
   601 GTCGCCATGA GCCTGATGAA CTATTACGCG CTGTTTGCCC TGATTATGGT
   651 ATTCGTGCTC GCATGGTTCT CCTTCGACAT CGGCTCGAtg gCGCGTTTCG
   701 AACAGGCTGC GTTGAACGAA gccaggacg aaaccgccc tTCAGACgCT
   751 ACCAAAGGTC GTGTTACGC ATTGATTATT CCCGTTTGG CCTTAATCGC
50 801 CTCAACGGTT TCCGCCATGA TCTACACCGG CGCGCAGGCA AGCGAAACCT
   851 TCAGCATTTT GGGGGCATT GAAAATACCG ACGTAAACAC TTCGCTGGTA
   901 TTCGGCGGCA CTTGCGGCGT GCTTGCCGTC GTCCTCTGCA CGTTCGGCAC
   951 GATTAAAACC ACCGATTATC CCAAAGCCGT GTGGCAGGCT GCGAAATCCA
1001 TGTTCCGCGC AATGCCATT TTAATCCTCG CCTGGCTCAT CAGTACGGTT
55 1051 GTCGGCGAAA TGCACACGGG CGACTACCTC TCCACGCTGG TTGCGGGCAA
   1101 CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCTGCTC GCCAGCGTGA
   1151 TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
   1201 ATTGCCGCCG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATcccGTG
   1251 TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGTTGCCCCA
60 1301 TCTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC
   1351 GACCACGTTA CCTCGCAACT GCCTTATGCC CTGACGGTTG CCGCCGCCGC
   1401 CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
   1451 TTGGCACGAC CGGTATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT
   1501 AAAAAACGCG CCGACGTTTG A

```

65 This encodes a protein having amino acid sequence <SEQ ID 696>:

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFVLVGGNPV  
 51 DGLTHLKDMV VGLAWADGDW SLGKPKILVF LILLGIFTSL LTYSGSNQAF  
 101 ADWAKRHIKN RCGAKMLTAC LVFVTFIDY FHS LAVGAIA RPVTDKFKVS  
 151 RAKLAYILDS TASPVCVLMV VSSWGASIIA TLAGLLVITYK ITEYTPMGTF  
 5 201 VAMSLMNYA LFALIMVFV AVFSFDIGSM ARFEQAALNE AQDETAASDA  
 251 TKGRVYALII PVLALIASTV SAMIYTGAQA SETFSILGAF ENTVDNTSLV  
 301 FGGTCGVLAV VLCTFGTIKT ADYPAVWQG AKSMFGAIAI LILAWLISTV  
 351 VGEMHTGDYL STL VAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP  
 401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDITIL SSTGARNHI  
 10 451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD  
 501 KKRADV\*

ORF26ng and ORF26-1 show 98.4% identity in 505 aa overlap:

15	orf26-1.pep	10 20 30 40 50 60	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV
	orf26ng	10 20 30 40 50 60	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV
20	orf26-1.pep	70 80 90 100 110 120	VGLAWSGDGWSLGKPKILVFLILGIFTSLTYSGSNQAFADWAKRHIKNRRGAKMLTAC
	orf26ng	70 80 90 100 110 120	VGLAWADGDWSLGKPKILVFLILGIFTSLTYSGSNQAFADWAKRHIKNRCGAKMLTAC
25	orf26-1.pep	130 140 150 160 170 180	LVFVTFIDYFHS LAVGAIA RPVTDKFKVSRTKLAYILDSTAAPMCVLMPVSSWGASIIA
	orf26ng	130 140 150 160 170 180	LVFVTFIDYFHS LAVGAIA RPVTDKFKVSRAKLAYILDSTASPMCVLMPVSSWGASIIA
30	orf26-1.pep	190 200 210 220 230 240	TLAGLLVITYKITEYTPMGTFVAMSLMNYALFALIMVFVAVWFSFDIGSMARFEQAALNE
	orf26ng	190 200 210 220 230 240	TLAGLLVITYKITEYTPMGTFVAMSLMNYALFALIMVFVAVWFSFDIGSMARFEQAALNE
35	orf26-1.pep	250 260 270 280 290 300	AHDETAVS DATKGRVYALII PVLALIASTVSAMIYTGAQASETFSILGAFENTVDNTSLV
	orf26ng	250 260 270 280 290 300	AQDETAAS DATKGRVYALII PVLALIASTVSAMIYTGAQASETFSILGAFENTVDNTSLV
45	orf26-1.pep	310 320 330 340 350 360	FGGTCGVLAVVLCTLGTIKTADYPAVWQGA KSMFGAIAI LILAWLISTVVGEMHTGDYL
	orf26ng	310 320 330 340 350 360	FGGTCGVLAVVLCTFGTIKTADYPAVWQGA KSMFGAIAI LILAWLISTVVGEMHTGDYL
50	orf26-1.pep	370 380 390 400 410 420	STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSA
	orf26ng	370 380 390 400 410 420	STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSA
55	orf26-1.pep	430 440 450 460 470 480	VMAGAVCGDHCSPISDITILSSTGARNHIDHVTSQLPYALTVA AAAASGYLALGLTKSA
	orf26ng	430 440 450 460 470 480	VMAGAVCGDHCSPISDITILSSTGARNHIDHVTSQLPYALTVA AAAASGYLALGLTKSA
60	orf26-1.pep	490 500	LLGFGTTGIVLAVLIFLLKDKKRANAX
	orf26ng	490 500	LLGFGTTGIVLAVLIFLLKDKKRADVX

In addition, ORF26 ng shows significant homology to a hypothetical *H.influenzae* protein:



-400-

sp|P44263|YF86\_HAEIN HYPOTHETICAL PROTEIN HI1586 >gi|1074850|pir||C64037  
 hypothetical  
 protein HI1586 - Haemophilus influenzae (strain Rd KW20) >gi|1574427 (U32832) H.  
 influenzae predicted coding region HI1586 [Haemophilus influenzae] Length = 519  
 Score = 538 bits (1370), Expect = e-152  
 Identities = 280/507 (55%), Positives = 346/507 (68%), Gaps = 7/507 (1%)

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40

Query: 1 MQLIDYSHSFFSVVPPFLALALAVITRRXXXXXXXXXXXXAFLVGGNPVDGLTHLKDMV 60  
 M+LID+S S +S+VP LA+ LA+ TRR L +L V  
 Sbjct: 14 MELIDFSSSVWSIVPALLAIILAIATRRVLVSLSAGIIIGSLMLSQIGSAFNVLVKNV 73

Query: 61 VGLAWADGDWSLKGPKILVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRCGAKMLTAC 120  
 V L +ADG+ + I++FL+LLG+ T+LLT SGSN+AFA+WA+ IK R GAK+L A  
 Sbjct: 74 VSLVYADGEIN-SNMNIVLFLLLGLVLTALLTVSGSNRAFAEWAQSRIGRRGAKLLAAS 132

Query: 121 LVFVTFIDDDYFHS LAVGAIA RPVTDKFKVSR AKLAYILDSTASPMCVLMPVSSWGASIIA 180  
 LVFVTFIDDDYFHS LAVGAIA RPVTD+FKVSR AKLAYILDSTA+PMCV+MPVSSWGA II  
 Sbjct: 133 LVFVTFIDDDYFHS LAVGAIA RPVTDKFKVSR AKLAYILDSTAAPMCVMPVSSWGAYIIT 192

Query: 181 TLAGLLVITYKITEYTPMGTFVAMSLMNYALFALIMVFFVAVWFSFDIGSMARFEQAALNE 240  
 + GLL TY ITEYTP+G FVAMS MN+YA+F++IMVF VA+FSFDI SM R E+ AL  
 Sbjct: 193 LIGLLATYSITEYTPIGAFVAMSSMNFYAI FSIIMVFFVAYFSFDIASMRHEKLALKN 252

Query: 241 AQDETAASDATKGRVYALIIPVLALIASTVSAMIYTGAQA----SETFSILGAFENTDVN 296  
 +D+ TKG+V LI+P+L LI +TVS MIYTGA+A + FS+LG FENT V  
 Sbjct: 253 TEDQLEEETGKQVRNLILPILVLIATVSMMIYTGAELAADGKVFVSLGTFENTVVG 312

Query: 297 TSLVFGGTCGVL--AVVLCFTFGTIKTADYPKAVWQGA KSMF GXXXXXXXXXXXXSTVVGEM 354  
 TSLV GG C ++ +++ + +Y ++ G KSM G + +VG+M  
 Sbjct: 313 TSLVVGGFCSIIISTLLIILDRQVSVPYVRSWIVGIKMSGAIAILFFAWTINKIVGDM 372

Query: 355 HTGDYLS TLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVEPALI 414  
 TG YLS+LV+GNI FLPVILF+L + MAF+TGTSWGTFGIMLP IAAAMA P L+  
 Sbjct: 373 QTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSWGTFGIMLP IAAAMAANAPELL 432

Query: 415 IPCMSAVMAGAVCGDHCSPISD TTILSSTGARNHIDHVT SQXXXXXXXXXXXXXXXXXXXX 474  
 +PC+SAVMAGAVCGDHCSP+SD TTILSSTGA+CNHIDHVT+Q  
 Sbjct: 433 LPCLSAVMAGAVCGDHCSPVSD TTILSSTGAKCNHIDHVTQLPYAATVATATSIGYIVV 492

Query: 475 XXXKSALLGFGTTGIVLAVLIFLLKDK 501  
 S L GF T + L V+IF +K +  
 Sbjct: 493 GFTYSGLAGFAATAVSLIVIIFAVKKR 519

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*,  
 45 and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 83

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 697>:

50  
 1 ..AAGCAATGGT ATGCCGACGN .AGTATCAAG ACGGAAATGG TTATGGTCAA  
 51 CGATGAGCCT GCCAAAATTC TGACTTGGGA TGAAAGCGGC CGATTACTCT  
 101 CGGAAGTGT CATTCCGCCAC CATCAACGCA ACGGGGTGGT TTTGGAGTGG  
 151 TATGAAGATG GTTCTAAAAA GAGCGAAGT. GTTTATCAGG ATGACAAGTT  
 201 GGTGAGGAAA ACCCAGTGGG ATAAGGATGG TTATTTAATC GAACCTGA

This corresponds to the amino acid sequence <SEQ ID 698; ORF27>:

55  
 1 ..KQWYADXSIK TEMVMVNDEP AKILTWDESG RLLSELSIRH HQRNGVVLEW  
 51 YEDGSKKSEX VYQDDKLVRK TQWDKDGILI EP\*

Further work revealed the complete nucleotide sequence <SEQ ID 699>:

60  
 1 ATGAAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC  
 51 GGCCGCTTTG CCGGCGCAGA CCTATTCTGT TTATTTTAAT CAGAACGGAA  
 101 AGCTGACGGC GACGATGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG  
 151 GTGGCGGGTA TTGCGCACGC GCAGGATTTT TATTATCCGT CGATGAAGAA

-401-

5  
10

```

201 ATATTCTGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC
251 CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA TGGTCAGAAA
301 AAAATGGCGG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA
351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAAATGGCT
401 TGAGTGAGGG TACGGGATAC CGCTATTACC GTAACGGCGG CAAGGAAAGC
451 GAAATCCAGT TTAAGCAAAA TAAGGCAAAC GGCATATGGA AGCAATGGTA
501 TGCCGACGGC AGTATCAAGA CGGAAATGGT TATGGTCAAC GATGAGCCTG
551 CCAAAATTCT GACTTGGGAT GAAAGCGGCC GATTACTCTC GGAAGTGTCT
601 ATCCGCCACC ATCAACGCAA CGGGGTGGTT TTGGAGTGGT ATGAAGATGG
651 TTCTAAAAAG AGCGAAGCTG TTTATCAGGA TGACAAGTTG GTCAGGAAAA
701 CCCAGTGGGA TAAGGATGGT TATTTAATCG AACCTGA

```

This corresponds to the amino acid sequence <SEQ ID 700; ORF27-1>:

15

```

1 MKKLSRIVFS TVLLGFSAAL PAQYYSVYFN QNGKLTATMS SAAYIRQYSV
51 VAGIAHAQDF YPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFNGQK
101 KMAGGFSK GK PDGEWVNWYP NGKKSAMVPY KNGLSEGTGY RYYRNGGKES
151 EIQFKQNKAN GVWKQWYADG SIKTEMVMVN DEPAKILTWD ESGRLLSELS
201 IRHHQRNGVV LEWYEDGSKK SEAVYQDDKL VRKTQWDKDG YLIEP*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF27 shows 91.5% identity over a 82aa overlap with an ORF (ORF27a) from strain A of *N. meningitidis*:

25

```

      orf27.pep                                10      20      30
                                         KQWYADXS I KTEMVMVNDEPAKILTWD E S G
                                         ||||| : |||||
      orf27a      LSEGTGXRYRNGGKESEI QFKQNKANGVWKQWYADGN I KTEMVMVNDEPAKILTWD E S G
                      140      150      160      170      180      190

      orf27.pep      40      50      60      70      80
      ||||| : || ||||| ||||| ||||| ||||| |||||
      orf27a      RLLSELSIRHHQRNGVVLEWYEDGSKKSEXYQDDKLVRKTQWDKDG YLIEP X
                      200      210      220      230      240

```

The complete length ORF27a nucleotide sequence <SEQ ID 701> is:

35

```

1 ATGAAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC
51 GGCCGCTTTG CCGGCGCAGA NCTATTCTGT TTATTTTAAAT CAGAACGGGA
101 AACTGACGGC GACGNTGCT TCTGCCGCNT ATATCAGGCA ATATAGTGTG
151 GCGGAGGGTA TTGCGCACGC GCAGGANTTT TANTATCCGT CGATGAAGAA
201 ATATTCCGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC
40 251 CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA NGTCAGAAA
301 AAAATGGCNG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA
351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAAATGGTT
401 TGAGTGAAGG TACGGGGTNN CGCTATTACC GTAACGGCGG CAAGGAAAGC
451 GAAATCCAGT TTAACAGAA TAAGGCAAAC GGCATATGGA AGCAATGGTA
501 TGCCGACGGC AATATCAAAA CGGAAATGGT TATGGTCAAT GATGAGCCTG
45 551 CCAAAATTCT GACATGGGAT GAAAGCGGTC GATTACTCTC GGAAGTGTCT
601 ATCCATCATC ATNAACGTAA TGGAGTAGTC TTAGAGTGGT ATGAAGATGG
651 TTCTAAAAAG ANTGAAGCTG TTTATCAGGA TGATAAGTTG GTCAGGAAAA
701 CCCAGTGGGA TAANGATGGT TATTTAATCG AACCTGA

```

This encodes a protein having amino acid sequence <SEQ ID 702>:

50

```

1 MKKLSRIVFS TVLLGFSAAL PAQXYSVYFN QNGKLTATXS SAAYIRQYSV
51 AEGIAHAQXF XYPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFXGQK
101 KMAGGFSK GK PDGEWVNWYP NGKKSAMVPY KNGLSEGTGX RYYRNGGKES
151 EIQFKQNKAN GVWKQWYADG NIKTEMVMVN DEPAKILTWD ESGRLLSELS
201 IHXXRNGVV LEWYEDGSKK XEAVYQDDKL VRKTQWDKDG YLIEP*

```

55 ORF27a and ORF27-1 show 94.7% identity in 245 aa overlap:

```

      10      20      30      40      50      60
orf27a.pep  MKKLSRIVFSTVLLGFSAALPAQXYSVYFNQNGKLTATXS SAAYIRQYSVAEGIAHAQXF

```

	orf27-1	:	
		MKKLSRIVFSTVLLGFSALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVVAGIAHAQDF	
		10 20 30 40 50 60	
5		70 80 90 100 110 120	
	orf27a.pep	XYPSPMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFXGQKKMAGGFSKGPDPGEVWNWYP	
	orf27-1	YYPSPMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSKGPDPGEVWNWYP	
		70 80 90 100 110 120	
10		130 140 150 160 170 180	
	orf27a.pep	NGKKSAVMPYKNGLSEGTGXRYRNGGKESEIQFKQNKANGVWKQWYADGNIKTEMVMVN	
		:	
	orf27-1	NGKKSAVMPYKNGLSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN	
		130 140 150 160 170 180	
15		190 200 210 220 230 240	
	orf27a.pep	DEPAKILTWDESGRLLSELSIHXXRNGVVLEWYEDGSKKXEAVYQDDKLVKRTQWDXDG	
		:	
20	orf27-1	DEPAKILTWDESGRLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVKRTQWDKDG	
		190 200 210 220 230 240	
25	orf27a.pep	YLIEPX	
	orf27-1	YLIEPX	

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF27 shows 96.3% identity over 82 aa overlap with a predicted ORF (ORF27ng) from

30 *N.gonorrhoeae*:

	orf27.pep	KQWYADXS IKTEMVMVNDEPAKILTWDESG	30
	orf27ng	LSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVNDEPAKILTWDESG	193
35	orf27.pep	RLLSELSIRHHQRNGVVLEWYEDGSKKSEXVYQDDKLVRKTQWDKDGYLIEP	82
	orf27ng	RLLSELSIRHHKRVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDGYLIEP	245

The complete length ORF27ng nucleotide sequence <SEQ ID 703> is:

40	1	ATGAAGAAAT	TATCTCGGAT	TGTATTTTCA	ATCGTACTGT	TGGGTTTTTC
	51	GGCCGCTTTG	CCGGCGCAGA	CCTATTCTGT	TTATTTTAAT	CAGAACGGGA
	101	AACTGACGGC	GACGATGTCT	TCTGCCGCTT	ATATCAGGCA	ATATAGTGTG
	151	GCGGCGGGTA	TCGCACACGC	GCAGGATTTT	TATTATCCGT	CGATGAAGAA
45	201	ATATTCCGAA	CCTTATATCG	TTGCTTCAAC	GCAAAATCAA	TC'TTTGTGTC
	251	CTACCCTGCA	AAACGGTATG	TTGATTTTGT	GGCATTTTAA	TGGTCAGAAA
	301	AAAATGGCGG	GGGGCTTCAG	CAAGGGTAAG	CCGGACGGGG	AATGGGTCAA
	351	CTGGTATCCG	AACGGTAAAA	AATCTGCGGT	TATGCCTTAT	AAAAATGGCT
50	401	TGAGTGAGGG	TACGGGATAC	CGTTATTACC	GTAAACGGCG	CAAGGAAAGC
	451	GAAATCCAGT	TTAAGCAAAA	TAAGGCGAAC	GGCGTATGGA	AGCAATGGTA
	501	TGCCGATGGA	AGTATCAAGA	CGGAAATGGT	TATGGTCAAC	GATGAGCCTG
	551	CCAAAATTCT	GACTTGGGAT	GAAAGCGGCC	GATTACTTTC	GGAACTGTCT
	601	ATCCGCCACC	ATAAACGCAA	CGGGGTGGTT	TTGGAGTGGT	ATGAAGATGG
	651	TTCTAAAAAG	AGCGAGGCTG	TTTATCAGGA	TGACAAGTTG	GTCAAGAAAA
	701	CCCAATTGGGA	TAAGGATGGT	TATTTAATCG	AACCTTGA	

This encodes a protein having amino acid sequence <SEQ ID 704>:

55           1   MKKLSRIVFS   IVLLGFSAAAL   PAQTYSVYFN   QNGKLTATMS   SAAYIRQYSV  
          51   AAGIAHAQDF   YYPSPMKKYSE   PYIVASTQIK   SFVPTLQNGM   LILWHFNGQK  
         101   KMAGGFSKGGK   PDGEWNNWYP   NGKKSAMVPP   KNGLSEGTFD   RYYRNGGKES  
         151   EIQFKQNKAN   GVWKQWYADG   SIKTEMVMVN   DEPAKITLWD   ESGRLLSELS  
         201   IRHHKRGNGV   LEWYEDGSKK   SEAVYQDDKL   VRKTDQWDKG   YLIEP\*

60 ORF27ng and ORF27-1 show 98.8% identity in 245 aa overlap:

orf27-1.pep MKKLSRVFSTVLLGFSAALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVVAGIAHAQDF

	orf27ng	MKKLSRIVFSIVLLGFSAAALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVAAGIAHAQDF       10                  20                  30                  40                  50                  60
5	orf27-1.pep	YYPSMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSGKGPDGEWVNWYP 70                  80                  90                  100                  110                  120
10	orf27ng	YYPSMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSGKGPDGEWVNWYP 70                  80                  90                  100                  110                  120
	orf27-1.pep	NGKKSAVMPYKNGLSEGTYRYRRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN 130                  140                  150                  160                  170                  180
15	orf27ng	NGKKSAVMPYKNGLSEGTYRYRRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN 130                  140                  150                  160                  170                  180
	orf27-1.pep	DEPAKILTWEDESGRLLSEL SIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKD G 190                  200                  210                  220                  230                  240
20	orf27ng	DEPAKILTWEDESGRLLSEL SIRHHKRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKD G 190                  200                  210                  220                  230                  240
25	orf27-1.pep	YLIEPX 
	orf27ng	YLIEPX

Based on this analysis, including the putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF27-1 (24.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 17A shows the results of affinity purification of the GST-fusion protein, and Figure 17B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result, confirming that ORF27-1 is a surface-exposed protein and a useful immunogen.

### Example 84

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 705>:

40	1	ATGAAATTTA	CCAAGCACCC	CGTCTGGGCA	ATGGCGTTCC	GCCCATTTTA
	51	TTCGCTGGCG	GCTCTGTACG	GCGCATTGTC	CGTATTGCTG	TGGGGTTTCG
	101	GCTACACGGG	AACGCACKAG	CTGTCCGGTT	TCTATTGGCA	CGCGCATGA <sub>g</sub>
	151	ATGATTTGGG	GTTATGCCGG	ACTGGTCGTC	ATCGCCTTCC	TGCTGACCGC
	201	CGTCGCCCAT	TGGACGGGGC	AGCCGCCCAC	CGCGGGCGGC	GTaTCTGGTC
45	251	GGCTTGACTA	TCTTTTGGCT	GGCTGCGCGG	ATTGCCGCCT	TTATCCCGGG
	301	TTGGGGTGGC	TCGGCAAGCG	GCATACTCGG	TACGCTGTTT	TTCTGGTACG
	351	GCGCGGTGTG	CATGGCTTTG	CCCGTTATCC	GTTGCGAGAA	TCACGCAAC
	401	TATGTT <sub>g</sub> CCG	TGTTCCGCGT	GTTCGTCTTG	GGCGGCACGC	ATGCGGCGTT
	451	CCACGTCCAG	CTGCACAACG	GCAACCTAGG	CGGACTCTTG	AGCGGATTGC
	501	AGTCGGGCTT	GGTGATG			

50 This corresponds to the amino acid sequence <SEQ ID 706; ORF47>:

```

1  MKETKHPVWA  MAFRPFYSLA  ALYGALSULL  WFGYGTGTHX  LSGFYWHAHE
51  MTWGYAGLVV  IAFLLTAVAT  WTGQPPTRG  VLVGLTIFWL  AARIAAFIPG
101 WGASASGILG  TLEFFWYGAV  MALPVIRSON  ORNYVAVFAL  FVLGGTHAAF

```

151 HVQLHNGNLG GLLSGLQSG L VM

Further work revealed the complete nucleotide sequence &lt;SEQ ID 707&gt;:

```

1  ATGAAATTTA CCAAGCACCC CGTCTGGGCA ATGGCGTTCC GCCCATTTTA
5  51  TTCGCTGGCG GCTCTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
151 ATGATTTGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
201 CGTCGCCACT TGGACGGGGC AGCCGCCAC GCGGGGCGGC GTTCTGGTCG
251 GCTTGACTAT CTTTTGGCTG GCTGCGCGGA TTGCCGCCTT TATCCCGGGT
301 TGGGGTGCGT CGGCAAGCGG CATACTCGGT ACCTGTGTTT TCTGGTACGG
10 351 CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TTCGCAGAAT CAACGCAACT
401 ATGTTGCCGT GTTCGCGCTG TTCGTCTTGG GCGGCACGCA TCGGCGGTTT
451 CACGTCCAGC TGCACAACGG CAACCTAGGC GGAATCTTGA GCGGATTGCA
501 GTCGGGCTTG GTGATGGTGT CGGGTTTTAT CGGTCTGATT GGTACGCGGA
551 TTATTTTCGT TTTTACGTCC AAACGCTTGA ATGTGCCGCA GATTCCAGT
15 601 CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTGCCCATGC TGAATGCCAT
651 GCTGATGGCG CACGGTGTGT TGGCTTGGCT GTCTGCCGTT TTTGCCCTTTG
701 CCGCAGGTGT GATTTTACC GTGCAGGTGT ACCGCTGGTG GTATAAATCC
751 GTGTTGAAAG AGCCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAA CCCGCTTTCC
20 851 TCAATCTGGG TGTGCATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
901 TTGGGCATGA TGGCGCGTAC CGCGCTTGGT CATACGGGCA ATCCGATTTA
951 TCCGCCGCC AAAGCCGTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
1001 CGCCCGTCCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
1051 AGCATCCGCA CCTCTTCGGT TTTGTTTGA CTCGCGCTTT TGGTGTATGC
25 1101 GTGGAAGTAT ATTCCTTGGC TGATTCGTCC GCGTTCGGAC GGCAGGCCCC
1151 GTTGA

```

This corresponds to the amino acid sequence &lt;SEQ ID 708; ORF47-1&gt;:

```

1  MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHE LSGFYWHAHE
30 51  MIWGYAGLVV IAFLLTAVAT WTGQPTRGG VLVLTIIFWL AARIAAFIPG
101 WGASASGILG TLFFWYGAVC MALPVIRSON QRNYVAVFAL FVLGGTHAAF
151 HVQLHNGNLG GLLSGLQSG VMVSGFIGLI GTRIIISFFTS KRLNVPQIPS
201 PKWVAQASLW LPMLTAMLMH HGVLAWLSAV FAFAGVIFT VQYRWWYK
251 VLKEPMLWIL FAGYLTGLG LIAVGASYFK PAFNLGVHL IGVGIGVLT
301 LGMMARTALG HTGNPIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
35 351 SIRTSSVLEA LALLVYAWKY IPWLIRPRSD GRPG*

```

Computer analysis of this amino acid sequence predicts a leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)ORF47 shows 99.4% identity over a 172aa overlap with an ORF (ORF47a) from strain A of *N.*40 *meningitidis*:

```

10      20      30      40      50      60
orf47.pep MKFTKHPVWAMAFRPFYSLAALYGALSVLLWFGYTGTHXLSGFYWHAHEMIWGYAGLVV
45 orf47a  MKFTKHPVWAMAFRPFYSLAALYGALSVLLWFGYTGTHXLSGFYWHAHEMIWGYAGLVV
10      20      30      40      50      60

70      80      90      100     110     120
orf47.pep IAFLLTAVATWTGQPTRGGVLVGLTIIFWLAARIAAFIPGWGASASGILGLTFFWYGAVC
50 orf47a  IAFLLTAVATWTGQPTRGGVLVGLTIIFWLAARIAAFIPGWGASASGILGLTFFWYGAVC
70      80      90      100     110     120

130     140     150     160     170
orf47.pep MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLSGLQSGGLVM
55 orf47a  MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLSGLQSGGLVMVSGFIGLI
130     140     150     160     170     180

60 orf47a  GTRIIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMHGVMPWLSAAFAAGVIFT
190     200     210     220     230     240

```

The complete length ORF47a nucleotide sequence <SEQ ID 709> is:

```

1  ATGAAATTTA CCAAGCACCC CGTTTGGGCA ATGGCGTTCC GCCCGTTTTA
51  TTACTGCGC GCTCTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
5   151  ATGATTTGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
201  CGTCGCCACT TGGACGGGGC AGCCGCCCAC GCGGGGCGGC GTTCTGGTCG
251  GCTTGACTAT CTTTTGGCTG GCTGCGCGGA TTGCCGCCTT TATCCCGGGT
301  TGGGGTGCGT CGGCAAGCGG CATACTCGGT ACGCTGTTT TCTGGTACGG
351  CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TTCGCAGAAT CAACGCAATT
10  401  ATGTTGCCGT GTTCGCGCTG TTCGTCTTGG GCGGTACGCA CGCGCGTTC
451  CACGTCCAGC TGCACAACGG CAACCTAGGC GGACTCTTGA GCGGATTGCA
501  GTCGGGCTTG GTGATGGTGT CGGGTTTTAT CGGTCTGATT GGTACGCGGA
551  TTATTTTCGT TTTTACGTCC AAACGGTTGA ATGTGCCGCA GATTCCCAGT
601  CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTGCCCATGC TGACCGCCAT
15  651  GCTGATGGCG CACGGCGTGA TGCCTTGGCT GTCGGCGGCT TTCGCGTTG
701  CGGCAGGTGT GATTTTACC GTGCAGGTGT ACCGCTGGTG GTATAAGCCT
751  GTGTTGAAAG AGCCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
801  CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAA CCCGCTTTC
851  TCAATCTGGG TGTGCATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
20  901  TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CATACGGGCA ATCCGATTTA
951  TCCGCCGCCC AAAGCCGTTT CCGTTGCGGT TTGGCTGATG ATGGCGGCAA
1001 CCGCCGTCCT TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
1051 AGCATACGCA CCTCTTCGGT TTTGTTTGCA CTCGCGCTT TGGTGATATG
1101 GTGGAAGTAT ATTCCTTGGC TGATTTCGTC GCGTTCGGAC GGCAGGCCCG
25  1151 GTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 710>:

```

1  MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHE LSGFYWHAHE
51  MIWGYAGLVV IAFLLTAVAT WTGQPTRGG VLVGLTIFWL AARIAAFIPG
30  101  WGASASGILG TLEFFYGAVC MALPVIRSON QRNYVAVFAL FVLGGTHAAF
151  HVQLHNGNLG GLLSGLQSL VMVSGFIGLI GTRIISFFTS KRLNVPQIPS
201  PKWVAQASLW LPMLTAMLMA HGVMPWLSAA FAFAAGVIFT VQVYRWYKYP
251  VLKEPMLWIL FAGYLFGLG LIAVGASYFK PAFLNLGVHL IGVGIGIVLT
301  LGM MARTALG HTGNPIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
351  SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*

```

ORF47a and ORF47-1 show 99.2% identity in 384 aa overlap:

```

10      20      30      40      50      60
orf47a.pep  MKFTKHPVWAMAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV
40  orf47-1  MKFTKHPVWAMAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV
      10      20      30      40      50      60
70      80      90      100     110     120
orf47a.pep  IAFLLTAVATWTGQPTRGGVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
45  orf47-1  IAFLLTAVATWTGQPTRGGVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
      70      80      90      100     110     120
130     140     150     160     170     180
orf47a.pep  MALPVIRSONQRNYVAVFALFVLGGTHAAHFVQLHNGNLGGLLSGLQSLVMVSGFIGLI
50  orf47-1  MALPVIRSONQRNYVAVFALFVLGGTHAAHFVQLHNGNLGGLLSGLQSLVMVSGFIGLI
      130     140     150     160     170     180
190     200     210     220     230     240
orf47a.pep  GTRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAHGVMPWLSAAFAFAAGVIFT
55  orf47-1  GTRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAHGVLAWSAVFAFAAGVIFT
      190     200     210     220     230     240
250     260     270     280     290     300
orf47a.pep  VQVYRWYKYPVLKEPMLWILFAGYLFGLGLIAVGASYFKPAFLNLGVHLIGVGIGIVLT
60  orf47-1  VQVYRWYKYPVLKEPMLWILFAGYLFGLGLIAVGASYFKPAFLNLGVHLIGVGIGIVLT
      250     260     270     280     290     300
310     320     330     340     350     360

```

5	orf47a.pep	LGM MARTALGHTGNPIYPPPKAVPVAFWLMMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA
	orf47-1	LGM MARTALGHTGNPIYPPPKAVPVAFWLMMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA
		310 320 330 340 350 360
		370 380
10	orf47a.pep	LALLVYAWKYIPWLIRPRSDGRPGX
	orf47-1	LALLVYAWKYIPWLIRPRSDGRPGX
		370 380

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF47 shows 97.1% identity over 172 aa overlap with a predicted ORF (ORF47ng) from *N.gonorrhoeae*:

15	ORF47	MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV	60
	ORF47ng	MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV	60
20	ORF47	IAFLLTAVATWTGQPPTRGGLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC	120
	ORF47ng	IAFLLTAVATWTGQPPTRGGLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAVC	120
25	ORF47	MALPVIRSQNRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGGLVM	172
	ORF47ng	MALPVIRSQNRNYVAVFAIFVLGGTHAAFHVQLHNGNLGGLLSGLQSGGLVMWVGFIGLI	180

The ORF47ng nucleotide sequence <SEQ ID 711> is predicted to encode a protein comprising amino acid sequence <SEQ ID 712>:

30	1	MKFTKHPVWA	MAFRPFYSLA	ALYGALSVLL	WGFGYTGTHE	LSGFYWHAHE
	51	MIWGYAGLVV	IAFLLTAVAT	WTGQPPTRG	VLVGLTAFWL	AARIAAFIPG
	101	WGAAASGILG	TLFFWYGAVC	MALPVIRSQN	RRNYVAVFAI	FVLGGTHAAF
	151	HVQLHNGNLG	GLLSGLQSG	VMWVGFIGLI	GMKIISFFTS	KRLKLPQIPS
	201	PKWVAHASLW	LPMLNAILMA	HRVMPWLSAA	FPFAAGVIFT	VQVYAGGITP
	251	IEETSCGSA	GICYRLGNSS	G		

The predicted leader peptide and transmembrane domains are identical (except for an Ile/Ala substitution at residue 87 and an Leu/Ile substitution at position 140) to sequences in the meningococcal protein (see also *Pseudomonas stutzeri* orf396, accession number e246540):

40	TM segments in ORF47ng			
	INTEGRAL	Likelihood = -5.63	Transmembrane	52 - 68
	INTEGRAL	Likelihood = -3.88	Transmembrane	169 - 185
	INTEGRAL	Likelihood = -3.08	Transmembrane	82 - 98
	INTEGRAL	Likelihood = -1.91	Transmembrane	134 - 150
	INTEGRAL	Likelihood = -1.44	Transmembrane	107 - 123
	INTEGRAL	Likelihood = -1.38	Transmembrane	227 - 243

Further work revealed the complete gonococcal DNA sequence <SEQ ID 713>:

45	1	ATGAAATTTA	CCAAACATCC	CGTCTGGGCA	ATGGCGTTCC	GCCCGTTTTA
	51	TTCACGCGG	GCACTGTACG	GCGCATTGTC	CGTATTGCTG	TGGGGTTTCG
50	101	GCTACACGGG	AACGCACGAG	CTGTCCGGTT	TCTATTGGCA	CGCGCATGAG
	151	ATGATTGGG	GTTATGCCGG	TCTCGTCGTC	ATCGCCTTCC	TGCTGACCGC
55	201	CGTCGCCACT	TGGACGGGAC	AGCGGCCAC	GAGGGCGGC	GTTCTGGTCG
	251	GCTTGACCGC	CTTTTGCGTG	GCTGCGCGGA	TTGCCGCTT	TATCCGGGT
	301	TGGGGTGCGG	CGGCAAGCGG	CATACTCGGT	ACGCTGTTTT	TCTGGTACGG
	351	CGCGGTGTGC	ATGGCTTTGC	CCGTATCCG	TtcgCAAAAC	CGGCGCAACT
	401	ATGtcgCCGT	ATTGCGAATA	TTTGTGCTGG	GCGGTACGCA	TGCGgcgTTC
	451	CACGtcAcg	tGCACAACGG	CAACCTAGGC	GGACTCTTGA	GCGGATTGCA
55	501	GTCGGGCTTG	GTTATGGTGT	CGGGCTTTAT	CGGCCTGATT	GGGATGAGGA
	551	TTATTTCGTT	TTTTACGTCC	AAACGGTTGA	ACGTGCCGCA	GATTCCCAGT
	601	CCGAAATGGG	TGGCGCAGGC	TCGCTGTGG	CTACCCATGC	TGACCGCCAT

-407-

5  
10  
15  
20

```

651 ACTGATGGCG CACGGCGTGA TGCCTTGGCT GTCGGCGGCT TTCGCGTTTG
701 CGGCGGGCGT GATTTTTACC GTACAGGTGT ACCGCTGGTG GTATAAACCC
751 GTATTGAAAG AACCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAAA CCTGCCTTCC
851 TCAATCTGGG CGTACATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
901 TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CACACGGGCA ATTCGATTTA
951 TCCGCCGCCC AAAGCCGTTT CCGTTGCGTT TTGGCTGATG ATGGCCGGCA
1001 CCGCCGTCGG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
1051 AGCATCCGCA CGTCTTCGGT TTTGTTTGCA CTCGCGCTGC TGGTGTATGC
1101 GTGGAATAC ATTCCGTGGC TGATCCGTCC GCGTTCGGAC GGCAGGCCCG
1151 GTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 714; ORF47ng-1>:

15  
20

```

1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFYGTGTHE LSGFYWHAHE
51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGG VLVGLTAFWL AARIAAFIPG
101 WGAASGILG TLFFWYGAVC MALPVIRSON RRNYVAVFAI FVLGGTHAAF
151 HVQLHNGNLG GLLSGLQSLG VMVSGFIGLI GMRIISFPTS KRLNVPQIPS
201 PKWVAQASLW LPMLTAILMA HGVMFWLSAA FAFAAGVIFT VQVYRWYKYP
251 VLKEPMLWIL FAGYLFGLG LIAVGASYFK PAFLNLGVHL IGVGIGVLT
301 LGMMARTALG HTGNSIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*

```

ORF47ng-1 and ORF47-1 show 97.4% identity in 384 aa overlap:

25  
30  
35  
40  
45  
50  
55  
60

```

          10      20      30      40      50      60
orf47-1.pep MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMIWGYAGLVV
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf47ng-1   MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMIWGYAGLVV
          10      20      30      40      50      60

          70      80      90     100     110     120
orf47-1.pep IAFLLTAVATWTGQPPTRGGVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf47ng-1   IAFLLTAVATWTGQPPTRGGVLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAVC
          70      80      90     100     110     120

          130     140     150     160     170     180
orf47-1.pep MALPVIRSONRNYVAVFALFVLGGTHAAAFHVQLHNGNLGGLLSGLQSLGVMVSGFIGLI
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf47ng-1   MALPVIRSONRNYVAVFAIFVLGGTHAAAFHVQLHNGNLGGLLSGLQSLGVMVSGFIGLI
          130     140     150     160     170     180

          190     200     210     220     230     240
orf47-1.pep GTRIISFPTSKRLNVPQIPSPKWVAQASLWLPMLTAMLAHGVLAWSAVFAFAAGVIFT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf47ng-1   GMRIISFPTSKRLNVPQIPSPKWVAQASLWLPMLTAILMAHGVMFWLSAAFAFAAGVIFT
          190     200     210     220     230     240

          250     260     270     280     290     300
orf47-1.pep VQVYRWYKYPVLKEPMLWILFAGYLFGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf47ng-1   VQVYRWYKYPVLKEPMLWILFAGYLFGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT
          250     260     270     280     290     300

          310     320     330     340     350     360
orf47-1.pep LGMMARTALGHTGNPIYPPPKAVPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf47ng-1   LGMMARTALGHTGNSIYPPPKAVPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA
          310     320     330     340     350     360

          370     380
orf47-1.pep LALLVYAWKYIPWLIRPRSDGRPGX
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf47ng-1   LALLVYAWKYIPWLIRPRSDGRPGX
          370     380

```

Furthermore, ORF47ng-1 shows significant homology to an ORF from *Pseudomonas stutzeri*:

65 gnl|PID|e246540 (273914) ORF396 protein [Pseudomonas stutzeri] Length = 396  
Score = 155 bits (389), Expect = 5e-37



-408-

Identities = 121/391 (30%), Positives = 169/391 (42%), Gaps = 21/391 (5%)

Query: 7 PVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFY-----WHAHEMIWGYAGLV 59  
 P+W +AFRPF+ +LY L++ LW +TG GF WH HEM++G+A +  
 5 Sbjct: 14 PIWRLAFRPFFLAGSLYALLAIPLVAAWTGLWP--GFQPTGGWLAWHRHEMLFGFAMAI 71

Query: 60 VIAFLLTAVATWTGQPPTRGGLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAV 119  
 V FLLTAV TWTGQ G LVGL A WLAAR+ ++ G AA L LF  
 10 Sbjct: 72 VAGFLLTAVQTWTGQTAPSGNRLVGLAAVWLAARL-GWLFGLPAAWLAPDLLFLVALVW 130

Query: 120 CMALPVIRSQNRNRYVAVFAIFVLGGTHAAFXXXXXXXXXXXXXXXXXXXXXMVSGFIGL 179  
 MA + + +RNY V + ++ G +V+ + L  
 15 Sbjct: 131 MMAQMLWAVRQKRNYPIVVVLSLMLGADVLIILTGLLQGNDAQRQGVLAGLWLVAALMAL 190

Query: 180 IGMRIISFFTSKRLNVPQIPSP-KWVAQASLWLPMLTAILMAHGV----MPWLSAAFAFA 234  
 IG R+I FFT + L P W+ A L + A+L A GV P L F A  
 20 Sbjct: 191 IGGRVIPFFTQRGLGKVDVAVKPVWLDVALLVGTGVIALHAFGVAMRPQPLLGLLFV-A 249

Query: 235 AGVIFTVQVYRWYKPVLPKEPMLWILFAGYLFTGLGLIAVGASYF-KPAFXXXXXXXXXXX 293  
 GV +++ RW+ K + K +LW L L+ + + +F A  
 25 Sbjct: 250 IGVGHLLRLMRWYDKGIWKVGLLWSLHVAMLWLVAAFGLALWHFGLLAQSSPSLHALSV 309

Query: 294 XXXXXXXXXXXMMARTALGHTGNSIYPPPKAVPVAFWLXXXXXXXXXXXXFSSGTAYTHSIR 353  
 M+AR LGHTG + P + AF L F S +  
 30 Sbjct: 310 GSMSGLILAMIARVTLGHTGRPLQLPAGIIG-AFVL---FNLGTAARVFLSVAPVPGGLW 365

Query: 354 TSSVLFALALLVYAWKYIPWLIRPRSDGRPG 384  
 ++V + LA +Y W+Y P L+ R DG PG  
 35 Sbjct: 366 LAAVCWTLAFALYVWRYAPMLVAARVDGHPG 396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 85

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 715>:

35 1 ..ATGCCGCTCTG AAGGTTTCAGA CGGCmTCGGT GyCGGGGAay CAGAAGyGGT  
 51 AGCGCATGCC CAATGAGACT TCGTGGGTTT TGAAGCGGGT GTTTTCCAAG  
 101 CGTCCCCAGT TGTGGTAACG GTATCCGGTG TCyAArGTCA GCTTGGGyGT  
 151 GATGTCGAAa CCGACACCGG CGATGACACC AAGACCyAmG CTGCTGATrC  
 40 201 TGTkGCTTTC GTGATAGGsA GGTtGyTGG kmksAsyTTG TAyrATwkkG  
 251 CCTssCwsTG kAGmGCCkTk CkyTGGTkkA swGrwArTAG TCGTGGTtTy  
 301 TkTtTyCACc GAATGAACyT GATGTTTAAC GTGTCCGTAG GCGACGCGCG  
 351 CGCCGATATA GGGTTTGAAT TTATCGTTGA GTTTGAAATC GTAAATGGCG  
 401 GACAAGCCGA GAGAAGAAAC GCGTGGAAG CTGCCGTTTC CCTGATGTTT  
 45 451 TGTTTGGGTT TCTTTGTAGT TGTGTTTAT CTCTTCAGTA ACTTTTTTAG  
 501 TAGAAGAATT ACTTTCTTTC CATTTTCTGT AACTGGCATA ATCTGCCGCT  
 551 ATTCTCCAGC CGCCGAAATC ..

This corresponds to the amino acid sequence <SEQ ID 716; ORF67>:

50 1 ..MPSEGSdGXG XGEXEXVAHA QXDFVGFEAG VFQASPVVVT VSGVXXQLGX  
 51 DVETDTGDDT KTXAADXVAF VIGRFxGXXL YXXAXXXAX XWXXXXSRGF  
 101 XXHRMNLfEN VSVGDARADI GFEFIVEFEI VNGGQAERRN GVEAAVSLMF  
 151 CLGFFVVVY LFSNfFSRRI TFFPFsvTGI ICRYSPAAEI ..

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF67 shows 51.8% identity over 199 aa overlap with a predicted ORF (ORF67ng) from

55 *N.gonorrhoeae*:

	orf67.pep					MPSEGS	SDGXGXGEXEXVAHAQXDFVGF	EAG	30
	orf67ng	TNFEIAVL	SGMTVRV	FYCARPA	PVNGGRL	KMPSEGS	SDGIGIGESEAVAHAQ	RGFVGF	146
		90	100	110	120	130	140		
5	orf67.pep	VFQASPV	VTVSGV	XXQLGX	DVETDT	GDDTKTX	AADXVAFVIGRFXGXLYXXAXXXAX		90
			: :		: :	: :	:	:	
	orf67ng	VFQASPV	VAVAGV	QGQAGR	DVYAHAR	HRAEAQA	AAAVAF	LIGVFL	206
								RMSSVR	
10	orf67.pep	XWXXXX	SRGFXX	HRMNL	MFNVSV	GDARAD	IGFEFIVEFEIVNGGQA	ERRNGVE	150
		:	:	: :	:	:	:	:	
	orf67ng	TRVGGK	STCYFF	SRIDAV	SDVSVG	DARTDIG	FEFVFE	FEIVNGGQA	266
								ERRNGVE	
	orf67.pep	CLGFFV	V-----	VVYLF	SNFFS	RRITFF	-PFSVTGI	ICRYSPA	190
			:	:	: :	:		:     :	
15	orf67ng	RLLVFY	VKLVA	AKSFIL	SFQLFY	VHGFIV	VFPVTGI	IRGDAP	326
								AAEVVAD	

The ORF67ng nucleotide sequence <SEQ ID 717> is predicted to encode a protein comprising amino acid sequence <SEQ ID 718>:

	1	MPSETVGSIV	NVGDES	VGFSPP	FPSIQHF	YRFHRI	HRIRLFR	PPGPMQL	
20	51	NRHSHGSGNL	GRGVWAT	VLSDKF	PCGQVRI	PACAGMT	NFEI	AVL	
	101	VFYCARPAPV	NGGRLKMPSE	GSDGIGIGES	EAVAHAQ	RGFVGFEAG	VQFQA		
	151	SPVVAVAGV	QGQAGR	DVYAHAR	HRAEAQA	AAAVAF	LIGVFL	RMSSVR	
	201	NCCVSITRVG	GKSTCYFFSR	IDAVSDVSVG	DARTDIGFEF	VVEFEIV	NGGQA	ERRNGVE	
	251	QAERRNGVEC	AVFLMFRL	LVFYV	KLVA	AKSFIL	SFQLFY	VHGFIV	
25	301	PVTGIIRGDA	PAAEVVAD	RHPGVDGM					
	351	IVGNAFGG	VG *						

Based on the presence of a several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 30 Example 86

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 719>

	1	ATGTTTGCTT	TTTGAAGC	CTTTTGTGTC	GAATACGGTT	ATGCGGCTGT	
	51	TTTTTTTGTA	TTGGTCATCT	GCGGTTTCGG	CGTGCCGATT	CCCGAGGATT	
35	101	TGACCTTGGT	AACAGGCGGC	GTGATTTCGG	GTATGGGTGA	TACCAATCCG	
	151	CATATTATGT	TTGCAGTCGG	TATGCTCGGC	GTATTGGTCG	GGGACGGCAT	
	201	CATGTTTCGCC	GCCGGACGAA	TTGGGGGGCA	GArArTCCTA	rGGTTCArAC	
	251	CTATTGCGsG	CATCATGACG	CCGrAACGTT	ATGAGCAGGT	TCAGGAAAAA	
	301	TTCGACAAAT	ACGGTAACTG	GGTCTTATTT	GTGCGCCGTT	TCCTGCCCGG	
	351	TTTGAGAACG	GCCGTATTTG	TTACAGCCGG	TATCAGCCGC	AAGGTTTCAT	
40	401	ACTTGCGTTT	TATCATTATG	GATGGACTGG	CCGCA...		

This corresponds to the amino acid sequence <SEQ ID 720; ORF78>:

	1	MFAFLEAFFV	EYGYAAVFFV	LVICGFGVPI	PEDLTLVTGG	VISGMGYTNP	
	51	HIMFAVGMLG	VLVGDGIMFA	AGRIWQXXL	XFXPIAXIMT	PXRYEQVQEK	
	101	FDKYGNWVLF	VARFLPGLRT	AVFVTAGISR	KVSYLRFIIM	DGLAA...	

### 45 Further work revealed the complete nucleotide sequence <SEQ ID 721>:

	1	ATGTTTGCTT	TTTGAAGC	CTTTTGTGTC	GAATACGGTT	ATGCGGCTGT	
	51	TTTTTTTGTA	TTGGTCATCT	GCGGTTTCGG	CGTGCCGATT	CCCGAGGATT	
	101	TGACCTTGGT	AACAGGCGGC	GTGATTTCGG	GTATGGGTGA	TACCAATCCG	
	151	CATATTATGT	TTGCAGTCGG	TATGCTCGGC	GTATTGGTCG	GGGACGGCAT	
50	201	CATGTTTCGCC	GCCGGACGAA	TTGGGGGGCA	GAAAATCCTA	AGGTTCAAAC	
	251	CTATTGCGCG	CATCATGACG	CCGAAACGTT	ATGAGCAGGT	TCAGGAAAAA	
	301	TTCGACAAAT	ACGGTAACTG	GGTCTTATTT	GTGCGCCGTT	TCCTGCCCGG	
	351	TTTGAGAACG	GCCGTATTTG	TTACAGCCGG	TATCAGCCGC	AAGGTTTCAT	
	401	ACTTGCGTTT	TATCATTATG	GATGGACTGG	CCGCACTGAT	TTCCGTCCCT	
55	451	ATTTGGATTT	ATCTGGCGGA	ATACGGTGCG	CACAACATCG	ATTGGCTGAT	

501 GGCAGAAAATG CACAGCCTGC AATCGGGTAT TTTGTTATC TTGGGTATAG  
 551 GTGCGACCGT TGTCGCTTGG ATTTGGTGGA AAAACGCCA ACGTATCCAG  
 601 TTTTACCGCA GCAAATTGAA AGAAAAGCGG GCGCAACGCA AAGCCGCCAA  
 651 GGCAGCCAAA AAAGCCGCGC AAAGCAAACA ATAA

5 This corresponds to the amino acid sequence <SEQ ID 722; ORF78-1>:

1 MFAFLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP  
 51 HIMFAVGMLG VLVDGIMFA AGRIWGQKIL RFKPIARIMT PKRYEQVQEK  
 101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFIIM DGLAALISVP  
 151 IWIYLGEYGA HNIDWLMAMK HSLQSGIFVI LGIGATVVAW IWWKKRQRIQ  
 201 FYRSKLKEKR AQRKAATAA KAAQSKQ\*

Computer analysis of this amino acid sequence predicts several transmembrane domains, and also gave the following results:

#### Homology with the dedA homologue of *H.influenzae* (accession number P45280)

ORF78 and the dedA homologue show 58% aa identity in 144aa overlap:

15 Orf78: 4 FLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGM--GYTNPHIMFAVGMLGV 61  
 FL FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+ N H+M V M+GV  
 DedA: 20 FLIGFFTEYGYWAVLFVLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHLMLLVSMIGV 79  
 20 Orf78: 62 LVGDGIMFAAGRIWGQXXLFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTA 121  
 L GD M+ GRI+G L F PI I+T R V+EKF +YGN VLFVARFLPGLR  
 DedA: 80 LAGDSMYWLGRIYGTILRFRPIRRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAP 139  
 Orf78: 122 VFVTAGISRKVSYLRFIIMDGLAA 145  
 +++ +GI+R+VSY+RF+++D AA  
 25 DedA: 140 IYMSGITRRVSYRVFLIDFCAA 163

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF78 shows 93.8% identity over a 145aa overlap with an ORF (ORF78a) from strain A of *N. meningitidis*:

30 orf78.pep 10 20 30 40 50 60  
 MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG  
 orf78a MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG  
 35 10 20 30 40 50 60  
 orf78.pep 70 80 90 100 110 120  
 VLVDGIMFAAGRIWGQXXLFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRT  
 orf78a VLVDGIMFAAGRIWGQKILFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT  
 40 70 80 90 100 110 120  
 orf78.pep 130 140  
 AVFVTAGISRKVSYLRFIIMDGLAA  
 orf78a AVFVTAGISRKVSYLRFIIMDGLAALISVPVWIYLGEYGAHNIDWLMAMKHSLSQSGIFIA  
 45 130 140 150 160 170 180

The complete length ORF78a nucleotide sequence <SEQ ID 723> is:

1 ATGTTTGCCC TTTTGGAGC CTTTTTTGTC GAATACGGCT ATGCGGCCGT  
 51 GTTTTTTCGT TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAGGATT  
 101 TGACCTTGGT AACAGGCGCG GTGATTTCGG GTATGGGTTA TACCAATCCG  
 151 CATATTATGT TTGCAGTCGG TATGCTCGGC GTATTGGTCG GGGACGGCAT  
 201 CATGTTTCGCC GCCGGACGCA TCTGGGGGCA GAAAATCCTC AAGTTCAAAC  
 251 CGATTGCGCG CATCATGACG CCGAAACGTT ACGCACAGGT TCAGGAAAAA  
 301 TTCGACAAAT ACGGCAACTG GGTGTTATTT GTCGCTCGTT TCCTGCCCGG  
 351 TTTGCGGACT GCCGTTTTTCG TTACCGCCGG CATCAGCCGC AAAGTATCGT  
 401 ATCTGCGCTT TCTGATTATG GACGGGCTTG CCGCGCTGAT TTCCGTGCCC  
 451 GTTTGGATTT ACTTGGGCGA GTACGGCGCG CACAACATCG ATTGGCTGAT

501 GGCAGAAATG CACAGCCTGC AATCCGGCAT CTCATCGCA TTGGGCGTGC  
 551 TGGCGGCGGC GCTGGCGTGG TTCTGGTGGC GCAAACGCCG ACATTATCAG  
 601 CTTTACCGCG CACAATTGAG CGAAAAACGC GCCAAACGCA AGGCGGAAAA  
 651 GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAA

5 This encodes a protein having amino acid sequence <SEQ ID 724>:

1 MFALLEAFFV EYGYAAVEFFV LVICGFGVPI PEDLTlVTGG VISGMGYTNP  
 51 HIMFAVGMLG VLVGDGIMFA AGRIWGQKIL KFKPIARIMT PKRYAQVQEK  
 101 FDKYGNWVLF VARFLPGLRT AVEFTAGISR KVSYLRFILIM DGLAALISVP  
 151 VWIYLGEYGA HNIDWLMAMK HSLQSGIFIA LGVLAAALAW FWRKRHHYQ  
 201 LYRAQLSEKR AKRKAKEAAK KAAQKQQ\*

ORF78a and ORF78-1 show 89.0% identity in 227 aa overlap:

		10	20	30	40	50	60
orf78a.pep		MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTlVTGGVISGMGYTNPHIMFAVGMLG					
15	orf78-1	MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTlVTGGVISGMGYTNPHIMFAVGMLG					
		10	20	30	40	50	60
		70	80	90	100	110	120
20	orf78a.pep	VLVDGIMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT					
	orf78-1	VLVDGIMFAAGRIWGQKILRFKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT					
		70	80	90	100	110	120
		130	140	150	160	170	180
25	orf78a.pep	AVEFTAGISRKVSYLRFILIMDGLAALISVPVWIYLGEYGAHNIDWLMAMKHSLSGIFIA					
	orf78-1	AVEFTAGISRKVSYLRFIIMDGLAALISVPIWIYLGEYGAHNIDWLMAMKHSLSGIFVI					
		130	140	150	160	170	180
30		190	200	210	220		
	orf78a.pep	LGVLAAALAWFWRKRHHYQLYRAQLSEKRAKRAKAAKAAQKQQX					
	orf78-1	LGIGATVVAWIWWKRQRIQFYRSKLKEKRAQRKAAKAAKAAQSKQX					
35		190	200	210	220		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF78 shows 97.4% identity over 38 aa overlap with a predicted ORF (ORF78ng) from *N. gonorrhoeae*:

40	orf78.pep	XXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTAVFVTAGISRKVSYLRF	137
	orf78ng	YPVLFVARFLPGLRTAVFVTAGISRKVSYLRF	32
	orf78.pep	IIMDGLAA	145
45	orf78ng	LIMDGLAALISVPVWIYLGEYGAHNIDWLMAMKHSLSGIFIALGVLAALAWFWRKR	92

The ORF78ng nucleotide sequence <SEQ ID 725> is predicted to encode a protein comprising amino acid sequence <SEQ ID 726>:

1 ..YPVLFVARFL PGLRTAVFVT AGISRKVSYL RFLIMDGLAA LISVPVWIYL  
 51 GEYGAHNIDW LMAKMHSLSQ GIFIALGVLA AALAWFWRK RRHYQLYRAQ  
 101 LSEKRAKRAK EKAAKKAAQK QQ\*

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 727>:

1 atgtttgccc tttTggaagc CTTTTTGTG GAAtacggCt atgcGGCCGT  
 51 GTTTTTTCGTT TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAAGATT  
 101 TGACCTTGGT AACGGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG  
 151 CATATTATGT TTGCGGTCGG TATGCTCGGC GTGTTGGCGG GCGACGGCGT  
 201 GATGTTTGCC GCCGGACGCA TCTGGGGGCA GAAAATCCTC AAGTTCAAAC  
 251 CGATTGCGCG CATCATGACG CCGAAACGTT ACGCGCAGGT TCAGGAAAAA  
 301 TTCGACAAAT ACGGCAACTG GGTCTGTTT GTCGCCCGTT TCCTGCCGGG

-412-

351 TTTGCGGACT GCCGTTTTTCG TTACCGCCCG CATCAGCCGC AAAGTATCGT  
 401 ATCTGCGCTT TCTGATTATG GACGGGCTGG CCGCGCTGAT TTCCGTGCCC  
 451 GTTTGGATTT ACTTGGGCGA GTACGGCGCG CACAACATCG ATTGGCTGAT  
 501 GGCGAAAATG CACAGCCTGC AATCGGGCAT CTTTCATCGCA TTGGGCGTGC  
 551 TGGCGGCGGC GCTGGCGTGG TTCTGGTGGC GCAAACGCCG ACATTATCAG  
 601 CTTTACGCGC CACAATTGAG CGAAAACGC GCCAAACGCA AGGCGGAAAA  
 651 GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAA

This corresponds to the amino acid sequence <SEQ ID 728; ORF78ng-1>:

1 MFALLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP  
 51 HIMFAVGMLG VLAGDGVMFA AGRIWGQKIL KFKPIARIMT PKRYAQVQEK  
 101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRLIM DGLAALISVP  
 151 VWIYLGEYGA HNIDWLMAKM HSLQSGIFIA LGVLAALAW FWRKRHHYQ  
 201 LYRAQLSEKR AKRKAEEKA KAAQKQ\*

ORF78ng-1 and ORF78-1 show 88.1% identity in 227 aa overlap:

15	orf78-1.pep	10	20	30	40	50	60
	orf78ng-1	10	20	30	40	50	60
20	orf78-1.pep	70	80	90	100	110	120
	orf78ng-1	70	80	90	100	110	120
25	orf78-1.pep	130	140	150	160	170	180
	orf78ng-1	130	140	150	160	170	180
30	orf78-1.pep	190	200	210	220		
	orf78ng-1	190	200	210	220		

Furthermore, orf78ng-1 shows homology to the dedA protein from *H. influenzae*:

sp|P45280|YG29\_HAEIN HYPOTHETICAL PROTEIN HI1629 >gi|1073983|pir||D64133 dedA  
 protein (dedA) homolog - Haemophilus influenzae (strain Rd KW20)  
 >gi|1574476 (U32836) dedA protein (dedA) [Haemophilus influenzae] Length = 212  
 Score = 223 bits (563), Expect = 7e-58  
 Identities = 108/182 (59%), Positives = 140/182 (76%), Gaps = 2/182 (1%)

Query: 5 LEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGM--GYTNPHIMFAVGMLGVL 62  
 L FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+ N H+M V M+GVL  
 Sbjct: 21 LIGFFTEYGYWAVLFVLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHLMLLVSMIGVL 80

Query: 63 AGDGVMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRTAV 122  
 AGD M+ GRI+G KIL+F+PI RI+T +R V+EKF +YGN VLFVARFLPGLR +  
 Sbjct: 81 AGDSCMYWLGRIYGTILRFRPIRRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAPI 140

Query: 123 FVTAGISRKVSYLRLIMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLSQSGIFIALG 182  
 ++ +GI+R+VS+Y+RF+++D AA+ISVP+WIIYLGE GA N+DWL ++ Q I+I +G  
 Sbjct: 141 YMVSGITRRVS YRVFLIDFCAAIISVPWIYLGEYGAHNIDWLMAKMHSLSQSGIFIALG 200

Query: 183 VL 184  
 L  
 Sbjct: 201 YL 202

Based on this analysis, including the presence of putative transmembrane domains, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 87

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 729>:

```

1 ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51 TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
101 AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCGCGTCGA
10 201 AGTGCATACC CACATCAACG ACAACGGCGT GATGCGGATG CGCGAAGTCG
251 AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC
301 AGCTATCATG TGATGTTTAT GGGTTTGAAA AAACAATTAA AAGAGGGCGA
351 TAAATTTCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC
401 AACTGGAAGT CAAATCGCG CCGATGCCGG CAATGAACCA C...
```

15 This corresponds to the amino acid sequence <SEQ ID 730; ORF79>:

```

1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDEA
51 KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGLK KQLKEGDKIP VTLKFKNAKA QTVQLEVKIA PMPAMNH...
```

Further work revealed the complete nucleotide sequence <SEQ ID 731>:

```

20 1 ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51 TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
101 AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCGCGTCGA
25 201 AGTGCATACC CACATCAACG ACAACGGCGT GATGCGGATG CGCGAAGTCG
251 AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC
301 AGCTATCATG TGATGTTTAT GGGTTTGAAA AAACAATTAA AAGAGGGCGA
351 TAAATTTCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC
401 AACTGGAAGT CAAATCGCG CCGATGCCGG CAATGAACCA CGGTCATCAC
451 CACGGCGAAG CGCATCAGCA CTAA
```

30 This corresponds to the amino acid sequence <SEQ ID 732; ORF79-1>:

```

1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDEA
51 KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGLK KQLKEGDKIP VTLKFKNAKA QTVQLEVKIA PMPAMNHGH
151 HGEAHQH*
```

35 Computer analysis of this amino acid sequence revealed a putative leader peptide and also gave the following results:

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF79 shows 94.6% identity over a 147aa overlap with an ORF (ORF79a) from strain A of *N.meningitidis*:

```

40          10      20      30      40      50      60
orf79.pep  MKKLLAAVMMAGLAGAVSAAGVHVEDGWARTTVEGMKIGGAFMKIHNDEAKQDFLLGGSS
          |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf79a     MKXLLAAVMMAGLAGAVSAAGIHVEDGWARTTVEGMKMGGAFMKIHNDEAKQDFLLGGSS
          10      20      30      40      50      60
45          70      80      90     100     110     120
orf79.pep  PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
          |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf79a     PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGXKKQLKXGDKIP
          70      80      90     100     110     120
50
```

```

              130      140
orf79.pep    VTLKFKNAKAQTVQLEVKIAPMPAMNH
              |||||
5  orf79a     VTLKFKNAKAQTVQLEVKTAPMSAMDHGHHHGEAHQH
              130      140      150

```

The complete length ORF79a nucleotide sequence <SEQ ID 733> is:

```

1  ATGAAANAAC TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
10 51 TTCCGCCGCC GGAATCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
101 AAGGTATGAA AATGGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCTGTGCGG ACCGCGTCGA
201 AGTGCATACC CATATCAATG ATAACGGTGT GATGCGGATG CGCGAAGTCG
251 AAGGCGGCGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCCGGC
301 AGCTATCATG TCATGTTTAT GGGTNTGAAA AAACAATTAA AAGANGGCGA
15 351 CAAGATTCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCA CAAACCGTCC
401 AACTGGAAGT CAAACCCGCG CCGATGTCGG CAATGGACCA CCGTCATCAC
451 CACGGCGAAG CGCATCAGCA CTA

```

This encodes a protein having amino acid sequence <SEQ ID 734>:

```

20 1  MKXLLAAVMM AGLAGAVSAA GIHVEDGWAR TTVEGMKMGG AFMKIHNDEA
51 51 KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGXK KQLKXGDKIP VTLKFKNAKA QTVQLEVKTA PMSAMDHGHH
151 HGEAHQH*

```

ORF79a and ORF79-1 show 94.9% identity in 157 aa overlap:

```

25      10      20      30      40      50      60
orf79a.pep MKXLLAAVMMAGLAGAVSAAGIHVEDGWAR TTVEGMKMGGAFMKIHNDEAKQDFLLGGSS
orf79-1     MKKLLAAVMMAGLAGAVSAAGVHVEDGWAR TTVEGMKIGGAFMKIHNDEAKQDFLLGGSS
              10      20      30      40      50      60
30      70      80      90      100     110     120
orf79a.pep PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGXKKQLKXGDKIP
orf79-1     PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
              70      80      90      100     110     120
35      130     140     150
orf79a.pep VTLKFKNAKAQTVQLEVKTAPMSAMDHGHHHGEAHQH
orf79-1     VTLKFKNAKAQTVQLEVKIAPMPAMNHGHHHGEAHQH
              130     140     150
40

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF79 shows 96.1% identity over 76 aa overlap with a predicted ORF (ORF79ng) from *N.gonorrhoeae*:

```

45 orf79.pep    FMKIHNDEAKQDFLLGGSSPVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGS 101
orf79ng                                INDNGVMRMREVKGVPLEAKSVTELKPGS 30
50 orf79.pep    YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKIAPMPAMNH 147
orf79ng        YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKTAPMSAMNHGHHHGEAHQH 86

```

An ORF79ng nucleotide sequence <SEQ ID 735> was predicted to encode a protein comprising amino acid sequence <SEQ ID 736>:

```

55 1  ..INDNGVMRM EVKGGVPLEA KSVTELKPGS YHVMFMGLKK QLKEGDKIPV
51 TLKFKNAKAQ TVQLEVKTAP MSAMNHGHHH GEAHQH*

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 737>:

```

1  ATGAAAAAAT TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51  TTccgccgCc GGagTccAtG TCGAggACGG CTGGGCGCGc accaCTGtcg
101 aaggtATgaa aatggGCGGC GCgttCATga aaATCCACAA CGACGaaGcc
5  151 atacaaGACt ttgtgcTCgg CGGaagcatg cccgttgccg accgcGTCGA
201 AGTGCAcata cacATCAACG ACAACGGCGT GATGCGTATG CGCGAAGTCA
251 AAGGCGGCGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCCGGC
301 AGCTATCACG TGATGTTTAT GGGTTTGAAG AAACAACCTGA AAGAGGGCGA
351 CAAGATTCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC
10 401 AACTGGAAGT CAAAACCGCG CCGATGTCGG CAATGAACCA CGGTCATCAC
451 CACGCGGAAG CGCATCAGCA CTAA

```

This corresponds to the amino acid sequence <SEQ ID 738; ORF79ng-1>:

```

1  MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKMG GAFMKIHND EA
51  IQDFVLGGSM PVADRVEVHT HINDNGVMRM REVKGGVPLE AKSVTELKPG
15 101 SYHVMFMGLK KQLKEGDKIP VTLKFKNAKA QTVQLEVKTA PMSAMNHGHH
151 HGEAHQH*

```

ORF79ng-1 and ORF79-1 show 95.5% identity in 157 aa overlap:

```

                10      20      30      40      50      60
orf79-1.pep  MKKLLAAVMMAGLAGAVSAAGVHVEDGWAR TTVEGMKIGGAFMKIHND EAKQDFLLGGSS
20  orf79ng-1  MKKLLAAVMMAGLAGAVSAAGVHVEDGWAR TTVEGMKMGGA FMKIHND EAIQDFVLGGSM
                10      20      30      40      50      60

                70      80      90      100     110     120
orf79-1.pep  PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
25  orf79ng-1  PVADRVEVHTHINDNGVMRMREVKG GVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
                70      80      90      100     110     120

                130     140     150
orf79-1.pep  VTLKFKNAKAQT VQLEVKIAPMPAMNHGHHHGEAHQH X
30  orf79ng-1  VTLKFKNAKAQT VQLEVKTAPMSAMNHGHHHGEAHQH X
                130     140     150

```

Furthermore, ORF79ng-1 shows significant homology to a protein from *Aquifex aeolicus*:

```

35  gi|2983695 (AE000731) putative protein [Aquifex aeolicus] Length = 151
    Score = 63.6 bits (152), Expect = 6e-10
    Identities = 38/114 (33%), Positives = 58/114 (50%), Gaps = 1/114 (0%)

Query: 24  VEDGWAR TTVEGMKMGGA FMKIHND EAIQDFVLGGSM PVADRVEVHTHINDNGVMRMREV 83
40  V+  W      G      M I N+  D+++G  +A RVE+H  + +N V +M
Sbjct: 27  VKHPWVMEPPPGPNTTMMGMII VNEGDEPDYLI GAKT DIAQRVELHKT VIENDVAKMVPQ 86

Query: 84  KGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIPVTLKFKNAKAQT VQLEV 137
45  +  +  +  K  E  K  YHVM +GLKK++KEGDK+ V L F+ +  TV+  V
Sbjct: 87  ER-IEIPPKGKVEFKHHGYHVMII GLKKRIKEGDKVKVELIFEKSGKITVEAPV 139

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF79-1 (15.6kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 18A shows the results of affinity purification of the His-fusion protein. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 18B) These experiments confirm that ORF79-1 is a surface-exposed protein, and that it is a useful immunogen.



**Example 88**

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 739>:

```

5      1  ATGACGGTAA CTGCGGCCGA AGGCGGCAAA GCTGCCAAGG CGTTAAAAAA
      51  ATATCTGATT ACGGGCATTT TGGTCTGGCT GCCGATTGCG GTAACGGTTT
     101  GGGTGGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACCTGCTG
     151  CCGAAGCAAT GGCGGCCGCA ATATGTTTTG GGGTTTAATA TCCCGGGGCT
     201  GGGCGTTATC GTTGCCATTG CCGTATTGTT TGTAAACCGA TTGTTTGCCG
     251  CCAACGTATT GGGTCGGCAG ATCCTCGCCG CGTGGGACAG CCTGTTGGGG
    10  301  CGGATTCCCG TTGTGAAATC CATCTATTCG AGTGTGAAAA AAGTATCCGA
     351  ATAcgTGCTG TCCGACAGCA GCCGTTCGTT TAAAACGCCG GTACTCGTGC
     401  CGTTTCCCCA GCCC GTATT TGGACGATyG CTTTCGTGTC AGGGCAGGTG
     451  TCGAATGCGG TTAAGGCCGC ATTGCCGAAs GACGGCGATT ATCTTCCGT
     501  TATGTGCCG ACCACGCCGA ATCCGACCGG CGGTTACTAT ATTATGGTAA
    15  551  AGAAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGGACGA AsCATTGAAA
     601  TATGTGATTT CGCTGGGTAT GGTATCCCT GACGACCTGC CCGTCAAAAC
     651  ATTGGCAsGA CCTATGCCGT CTGAAAAGGC GGATTGCCC GAACAACAAT
     701  AA

```

This corresponds to the amino acid sequence <SEQ ID 740; ORF98>:

```

20      1  MTVTAAEGGK AAKALKKYL I TGILVWLPIA VTVWVVSIV SASDQLVNLL
      51  PKQWRPQYVL GFNIPGLGVI VAI AVLFTG LFAANVLGRQ ILAAWDSLGL
     101  RIPVVKSIYS SVKKVSEYVL SDSSRSFKTP VLVFPQPGI WTIAFVSGQV
     151  SNAVKAALPX DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEXLK
     201  YVISLGMVIP DDL PVKTLAX PMPSEKADLP EQQ*

```

25 Further work revealed the complete nucleotide sequence <SEQ ID 741>:

```

30      1  ATGACGGAAC nTGCGGCCGA AGGCGGCAAA GCTGCCAArG CGTTAAAAAA
      51  ATATCTGATT ACGGGCATTT TGGTCTGGCT GCCGATTGCG GTAACGGTTT
     101  GGGTGGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACCTGCTG
     151  CCGAAGCAAT GGCGGCCGCA ATATGTTTTG GGGTTTAATA TCCCGGGGCT
     201  GGGCGTTATC GTTGCCATTG CCGTATTGTT TGTAAACCGA TTGTTTGCCG
     251  CCAACGTATT GGGTCGGCAG ATCCTCGCCG CGTGGGACAG CCTGTTGGGG
     301  CGGATTCCCG TTGTGAAATC CATCTATTCG AGTGTGAAAA AAGTATCCGA
     351  ATCGCTGCTG TCCGACAGCA GCCGTTCGTT TAAAACGCCG GTACTCGTGC
     401  CGTTTCCCCA GCCC GTATT TGGACGATTG CTTTCGTGTC AGGGCAGGTG
    35  451  TCGAATGCGG TTAAGGCCGC ATTGCCGAAG GACGGCGATT ATCTTCCGT
     501  GTATGTTCCG ACCACGCCGA ATCCGACCGG CGGTTACTAT ATTATGGTAA
     551  AGAAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGGACGA AGCATTGAAA
     601  TATGTGATTT CGCTGGGTAT GGTATCCCT GACGACCTGC CCGTCAAAAC
     651  ATTGGCAGGA CCTATGCCGT CTGAAAAGGC GGATTGCCC GAACAACAAT
    40  701  AA

```

This corresponds to the amino acid sequence <SEQ ID 742; ORF98-1>:

```

45      1  MTEXAAEGGK AAKALKKYL I TGILVWLPIA VTVWVVSIV SASDQLVNLL
      51  PKQWRPQYVL GFNIPGLGVI VAI AVLFTG LFAANVLGRQ ILAAWDSLGL
     101  RIPVVKSIYS SVKKVSESL SDSSRSFKTP VLVFPQPGI WTIAFVSGQV
     151  SNAVKAALPK DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEALK
     201  YVISLGMVIP DDL PVKTLAG PMPSEKADLP EQQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF98 shows 96.1% identity over a 233aa overlap with an ORF (ORF98a) from strain A of *N. meningitidis*:

```

55      orf98.pep      10      20      30      40      50      60
      orf98a      MTVTAAEGGKA AAKALKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL
      || |||||
      MTEPAAEGGKA AAKALKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL
      10      20      30      40      50      60

```

		70	80	90	100	110	120
	orf98.pep	GFNIPGLGVIVAI	AVLFTGLFAANVLGRQ	ILAAWDSLLGRIPVVK	SIYSSVKKVSEYVL		
5	orf98a	GFNIPGLGVIVAI	AVLFTGLFAANVLGRQ	ILAAWDSLLGRIPVVK	SIYSSVKKVSXSL		
		70	80	90	100	110	120
		130	140	150	160	170	180
	orf98.pep	SDSSRSFKTPVLV	PFPPQGIWTIAFVSGQV	SNVKAALPKDGDYLSV	YVPTTPNPTGGYY		
10	orf98a	SDSSRSFKTPVLV	PFPPQGIWTIAFVSGQV	SNVKAALPKDGDYLSV	YVPTTPNPTGGYY		
		130	140	150	160	170	180
		190	200	210	220	230	
	orf98.pep	IMVKKSDVRELDMS	VDEALKYVISLGMVIPDD	LPVKTLAXPMPSEKADL	PEQQX		
15	orf98a	IMVKKSDVRELDMS	VDEALKYVISLGMVIPDD	LPVKTLAXPMPSEKADL	PEQQX		
		190	200	210	220	230	

The complete length ORF98a nucleotide sequence <SEQ ID 743> is:

20	1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA
	51	ATATCTGATT	ACGGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTACCGGTTT
	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ATCAGCTCGT	CAACCTGCTG
	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTTG	GGGTTTAATA	TCCCGGGGCT
	201	GGGCGTTATC	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTATTTGCCG
25	251	CAAACGTATT	GGGCCGGCAG	ATTCTTGCCG	CGTGGGACAG	CTTGTGGGG
	301	CGGATTCCGG	TTGTGAAGTC	CATCTATTCG	AGTGTGAAAA	AAGTATCCGA
	351	NTCGTTGCTG	TCCGACAGCA	GCCGTTCTGT	TAAACACCA	GTACTCGTGC
	401	CGTTTCCCCA	ATCGGTATT	TGGACAATCG	CATTCGTGTC	CGGTCAGGTG
	451	TCGAATGCGG	TTAAGGCCGC	ATTGCCGAAG	GACGGCGATT	ATCTTCCGT
30	501	GTATGTTCCG	ACCACGCCGA	ATCCGACCGG	CGGTACTAT	ATTATGGTAA
	551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCGTTGAAA
	601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAAC
	651	ATTGGCAGGA	CCTATGCCGT	CTGAAAAGGC	GGATTGCCC	GAACAACAAT
	701	AA				

35 This encodes a protein having amino acid sequence <SEQ ID 744>:

	1	MTEPAAEGGK	AAKALKKYLI	TGILVWLPIA	VTWVVSIV	SASDQLVNLL
	51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTG	LFAANVLGRQ	ILAAWDSLLG
	101	RIPVVKSIYS	SVKKVSXSL	SDSSRSFKTP	VLVFPQSGI	WTIAFVSGQV
	151	SNVKAALPK	DGDYLSVYVP	TPNPTGGYY	IMVKKSDVRE	LDMSVDEALK
40	201	YVISLGMVIP	DDLVPKTLA	GMPSEKADLP	EQQ*	

ORF98a and ORF98-1 show 98.7% identity in 233 aa overlap:

		10	20	30	40	50	60
	orf98a.pep	MTEPAAEGGKA	AKALKKYLITGILVWLPIA	VTWVVSIVSASDQLVNLL	PKQWRPQYVL		
45	orf98-1	MTEXAAEGGKA	AKALKKYLITGILVWLPIA	VTWVVSIVSASDQLVNLL	PKQWRPQYVL		
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf98a.pep	GFNIPGLGVIVAI	AVLFTGLFAANVLGRQ	ILAAWDSLLGRIPVVK	SIYSSVKKVSXSL		
50	orf98-1	GFNIPGLGVIVAI	AVLFTGLFAANVLGRQ	ILAAWDSLLGRIPVVK	SIYSSVKKVSESLL		
		70	80	90	100	110	120
		130	140	150	160	170	180
	orf98a.pep	SDSSRSFKTPVLV	PFPPQGIWTIAFVSGQV	SNVKAALPKDGDYLSV	YVPTTPNPTGGYY		
55	orf98-1	SDSSRSFKTPVLV	PFPPQGIWTIAFVSGQV	SNVKAALPKDGDYLSV	YVPTTPNPTGGYY		
		130	140	150	160	170	180
		190	200	210	220	230	
	orf98a.pep	IMVKKSDVRELDMS	VDEALKYVISLGMVIPDD	LPVKTLA	GMPSEKADLPEQQX		
60	orf98-1	IMVKKSDVRELDMS	VDEALKYVISLGMVIPDD	LPVKTLA	GMPSEKADLPEQQX		
		190	200	210	220	230	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF98 shows 95.3% identity over a 233 aa overlap with a predicted ORF (ORF98ng) from *N.gonorrhoeae*:

5	orf98.pep	MTVTAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	60
	orf98ng	MTEPAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	60
10	orf98.pep	GFNIPGLGVIVAVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSEYVL	120
	orf98ng	GFNIPGLGVIVAVLFTGLFAANVLGRQILAAWDSLLXRIIPVVKSIYSSVKKVSESL	120
15	orf98.pep	SDSSRSFKTPVLVPPFPQGIWTIAFVSGQVSNVKAALPXDGDYLSVYVPTTPNPTGGYY	180
	orf98ng	SDSSRSFKTPVLVPPFPQSGIWTIAFVSGQVSNVKAALPQDGDYLSVYVPTTPNPTGGYY	180
20	orf98.pep	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPKTLAXPMPSEKADLPEQQ	233
	orf98ng	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPKTLAGPMPPEKAELPEQQ	233

The complete length ORF98ng nucleotide sequence <SEQ ID 745> is predicted to encode a protein having amino acid sequence <SEQ ID 746>:

1	MTEPAAEGGK	AAKALKKKYLI	TGILVWLPIA	VTWVVSIV	SASDQLVNLL
51	PKQWRPQYVL	GFNIPGLGVI	VAVLFTGL	LFAANVLGRQ	ILAAWDSLLX
101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTP	VLVPPFPQSGI	WTIAFVSGQV
151	SNAVKAALPQ	DGDYLSVYVP	TPNPTGGYY	IMVKKSDVRE	LDMSVDEALK
201	YVISLGMVIP	DDLVPKTLAG	PMPPEKAELP	EQQ*	

Further work revealed the complete nucleotide sequence <SEQ ID 747>:

1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA
51	ATATCTGATT	ACAGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT
101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ACCAGCTTGT	CAACCTGCTG
151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTTG	GGGTTTAATA	TCCCCGGGCT
201	CGGCGTTATT	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTATTGCGCG
251	CAAACGTGTT	GGGCGGCCAG	ATTCTTGCCG	CGTGGGACAG	CCTGTTgggg
301	cggattccgg	TTGTCAAATC	CATCTATTCG	AGTGTGAAAA	AAGTATCCGA
351	ATCGCTGCTG	TCCGACAGCA	GCCGTTTCGT	TAAACGCCG	GTACTCGTGC
401	CGTTTCCCCA	ATCGGGTATT	TGGACAATCG	CATTCGTGTC	CGGTCAGGTG
451	TGCAATGCCG	TTAAGGCCG	ATTGCCGCG	GATGGCGATT	ATCTTCCGT
501	GTATGTCCCG	ACCACGCCCA	ACCCGACCGG	CGGTTACTAT	ATTATGGTAA
551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCGTTGAAA
601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAAC
651	ATTGGCAGGA	CCTATGCCGC	CTGAAAAGGC	GGAGTTGCCC	GAACAACAAT
701	AA				

This corresponds to the amino acid sequence <SEQ ID 748; ORF98ng-1>:

1	MTEPAAEGGK	AAKALKKKYLI	TGILVWLPIA	VTWVVSIV	SASDQLVNLL
51	PKQWRPQYVL	GFNIPGLGVI	VAVLFTGL	LFAANVLGRQ	ILAAWDSLLG
101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTP	VLVPPFPQSGI	WTIAFVSGQV
151	SNAVKAALPQ	DGDYLSVYVP	TPNPTGGYY	IMVKKSDVRE	LDMSVDEALK
201	YVISLGMVIP	DDLVPKTLAG	PMPPEKAELP	EQQ*	

ORF98ng-1 and ORF98-1 show 97.9% identity in 233 aa overlap:

50	orf98-1.pep	MTEXAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	60
	orf98ng-1	MTEPAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	60
55	orf98-1.pep	GFNIPGLGVIVAVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSESL	120

-419-

orf98ng-1		GFNIPGLGVIVAVIAVLFTVGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSESL				
		70	80	90	100	110 120
5	orf98-1.pep	SDSSRSFKTPVLVPPFPQGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY	130	140	150	160 170 180
	orf98ng-1	SDSSRSFKTPVLVPPFPQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY	130	140	150	160 170 180
10	orf98-1.pep	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPVKTLAGPMPSEKADLPEQQX	190	200	210	220 230
	orf98ng-1	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPVKTLAGPMPPEKAELPEQQX	190	200	210	220 230

- 15 Based on this analysis, including the fact that the putative transmembrane domains in the gonococcal protein are identical to the sequences in the meningococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 89

- 20 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 749>:

25	1	ATgAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CCGTCGGACT
	51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
30	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CTTGTTTAAA	TTCATTATCG	G <sub>3</sub> G <sub>9</sub> TACTCA
35	201	ATATCCCCGA	AAAGATGCAG	CGTTTCGGTT	CGGCnCGTAA	AGGCCkCAAG
	251	ssCGsGCTTG	CCTTGAACAA	GGCGGGTTTG	GCGTATTTTG	AAGGGCGTTT
40	301	TGAAAAGGCG	GAAC TAGAAG	CCTCACGCGT	GTTGGTCAAC	AAAgTAGGCC
	351	GaGAGACAAC	CGGACTTTGG	CATTGATGCT	GrGCGCGCAC	GCCGCCGGAC
45	401	AGATGGAAAA	CATCGAsTG	CGCGACCGTT	ATCTTGCGGA	AATCGCCAAA
	451	CTGCCGGA	AACAGCAGCT	TTCCCGTTAT	CTTTTGTGG	CGGAATCGGC
50	501	GTTGAACCGG	CGCGATTACG	AAGCGGCGGA	AGCCAATCTT	CATGCGGCGG
	551	CGAAGATGAA	TGCCAACCTT	ACGCGCCTCG	TGCGTCTGCA	.ATTCGTTAC
55	601	GCTTTCGACA	GGGGCGACGC	GTTGCAGGTT	CTGGCAAAA	CCGAAAAACT
	651	TTCCAAGGCG	GGCGCGTTGG	GCAAATCGGA	AATGGAACGG	TATCAAAATT
60	701	GGGCATATCC	GTCGCCAGCT	GGCGGATGCT	GCCGATGCCG	CCGCTTTGAA
	751	AACCTGCCTG	AAGCGGATTC	CCGACAGCCT	CAAAAACGGG	GAATTGAGCG
65	801	TATCGGTTGC	GGAAAGTAC	GAACGTTTGG	GACTGTATGC	CGATGCGGTC
	851	AAATGGGTCA	AACAGCATT	TCCGCAaAC	CGCCGCCCGG	AGCTTTTGG
70	901	AGCCTTTGTC	GAAAGCGTGC	GCTTTTGGG	CGAGCGCGAA	CAGCAGAAAG
	951	CCATCGATTT	TGCCGATGCT	TGGCTGAAAG	AACAGCCCGA	TAACGCGCTT
75	1001	CTGCTGATGT	ATCTCGGTGC	GCTCGCCTTC	GGCCGCAAAC	TTTGGGGCAA
	1051	GGCAAAAGGC	TACCTTGAAG	CGAGCATTGC	ATTAAAGCCG	AGTATTTCCG
80	1101	CGCGTTTGGT	TCTAACAAAG	GTTTTCGACG	AAATCGGAGA	ACCGCAGAAG
	1151	GCGGAGGCCG	AC...			

- 45 This corresponds to the amino acid sequence <SEQ ID 750; ORF100>:

50	1	MKTVVWIVVL	FAAAVGLALA	SGIYTGdVYI	VLGQTMLRIN	LHAFVLGSLI
	51	AVVVWYFLFK	FIIGVLNIPE	KMQRFGSARK	GKXXLALNK	AGLAYFEGRF
55	101	EKAELEASRV	LVNKVGRDNR	TLALMLXAHA	AGQMenIXXR	DRYLAEIAKL
	151	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH	AAAKMNANLT	RLVRLXIRYA
60	201	FDRGDALQVL	AKTEKLSKAG	ALGKSEMERY	QnWAYRRQLA	DAADAAALKT
	251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWVKQHYP	XNRRPELLEA
65	301	FVESVRFLGE	REQQKAIDFA	DAWLKEQPDN	ALLMYLGRRL	AFGRKLWGKA
	351	KGYLEASIAL	KPSISARLVL	TKVFDEIGEP	QKAEAH...	

Further work revealed the complete nucleotide sequence <SEQ ID 751>:

55	1	ATGAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CCGTCGGACT
	51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
60	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CTTGTTTAAA	TTCATTATCG	GCGTACTCAA

-420-

201 TATCCCCGAA AAGATGCAGC GTTTCGGTTC GCGCGTAAA GGCCGCAAGG  
 251 CCGCGCTTGC CTTGAACAAG GCGGGTTTGG CGTATTTTGA AGGGCGTTTT  
 301 GAAAAGGCGG AACTAGAAGC CTCACGCGTG TTGGTCAACA AAGAGGCCGG  
 351 AGACAACCGG ACTTTGGCAT TGATGCTGGG CGCGCACGCC GCCGACAGA  
 5 TGGAAAAACAT CGAGCTGCGC GACCGTTATC TTGCGGAAAT CGCCAAACTG  
 451 CCGAAAAAAC AGCAGCTTTC CCGTTATCTT TTGTTGGCGG AATCGGCGTT  
 501 GAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA  
 551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCT  
 601 TTCGACAGGG GCGACGCGTT GCAGGTTCTG GCAAAAACCG AAAAAGCTTC  
 10 CAAGCGGGC GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATTGGG  
 701 CATAACGCGC CCAGCTGGCG GATGCTGCCG ATGCCGCCGC TTTGAAAACC  
 751 TGCCTGAAGC GGATTCCCGA CAGCCTCAA AACGGGAAT TGAGCGTATC  
 801 GGTTCGGGAA AAGTACGAAC GTTGGGACT GTATGCCGAT GCGGTCAAAT  
 851 GGGTCAAACA GCATTATCCG CACAACGCC GCCCGAGCT TTTGGAAGCC  
 15 TTTGTCGAAA GCGTGCGCTT TTTGGGCGAG CGCGAACAGC AGAAAGCCAT  
 951 CGATTTTGCC GATGCTTGGC TGAAAGAACA GCCCGATAAC GCGCTTCTGC  
 1001 TGATGTATCT CGGTCGGCTC GCCTACGCC GCAAACCTTG GGGCAAGGCA  
 1051 AAAGGCTACC TTGAAGCGAG CATTGCATTA AAGCCGAGTA TTCCGCGCG  
 1101 TTTGGTTCTA GCAAAGGTT TCGACGAAAT CGGAGAACC CAGAAGCGG  
 20 AGGCGCAGCG CAACTTGGT TTGGAAGCCG TCTCCGATGA CGAACGTCAC  
 1151  
 1201 GCAGCGTTAG AGCAGCATAG CTGA

This corresponds to the amino acid sequence <SEQ ID 752; ORF100-1>:

1 MKTVVWIVVL FAAVGLALA SGIYTGdVYI VLgQTMlRIN LHAfVLGSli  
 51 AVVVWYFLFK FIIGVLNIPE KMQRFGSARK GRKAALALNK AGLAYFEGRF  
 25 EKAELEASRV LVNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAELAKL  
 151 PEKQQLSRYL LLAESALNRR DYEAEEANLH AAKMNANLT RLVRQLRYA  
 201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQLA DAADAAALKT  
 251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWVKQHYP HNRRLPELLEA  
 301 FVESVRFLGE REQQAIDFA DAWLKEQPDN ALLLMYLGR LAYGRKLWGKA  
 351 KGYLEASIAL KPSISARLVL AKVFDEIGEP QKAEAQRLNV LEAVSDDERH  
 401 AALEQHS\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF100 shows 93.5% identity over a 386aa overlap with an ORF (ORF100a) from strain A of *N.*

35 *meningitidis*:

		10	20	30	40	50	60
orf100.pep		MKT	VVW	IVVL	FAAA	VGLALA	SGIYTGdVYI
orf100a		MKT	VVW	IVVL	FAAA	XGLALA	SGIXTGdVYI
		10	20	30	40	50	60
		70	80	90	100	110	120
orf100.pep		FIIG	VLNI	PEKM	QRFG	SARKGX	KXXLALNKAG
orf100a		FIIG	VLNX	PEKM	QRFG	SARKGR	KAALALNKAG
		70	80	90	100	110	120
		130	140	150	160	170	180
orf100.pep		TLAL	MLXA	HAAG	QMEN	IXXR	DRYLAELAKL
orf100a		TLAL	MLGA	HAAG	QMEN	IELRDR	YLAELAKL
		130	140	150	160	170	180
		190	200	210	220	230	240
orf100.pep		AAAK	MNAN	LTRL	VLRL	XIRYAF	DRGDALQVLAK
orf100a		AAAK	MNAN	LTRL	VLRL	QLRYAF	DRGDALQVLAK
		190	200	210	220	230	240
		250	260	270	280	290	300
orf100.pep		DAAD	AAALK	TCLK	RIPD	SLKNG	ELSVSVAEKYER
orf100a		DAAD	AAALK	TCLK	RIPD	SLKNG	ELSVSVAEKYER
		250	260	270	280	290	300

-421-

		310	320	330	340	350	360
	orf100.pep	FVESVRFLGEREQQKAIDFADAWLKEQPDNALLMYLGR	LAFGRKLWGKAKGYLEASIAL				
5	orf100a	FVESVRFLGERDQQKAIDFADAWLKEQPDNALLXYLGR	LAYGRKLWGKAKGYLEASIAL				
		310	320	330	340	350	360
		370	380				
	orf100.pep	KPSISARLVLT	KVDFEIGEPQKAEAH				
10	orf100a	KPSISARLVLA	KVDFETGEPQKAEQRNLVLA	SVAEENRPSAETHX			
		370	380	390	400		

The complete length ORF100a nucleotide sequence <SEQ ID 753> is:

	1	ATGAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CNNTCGGGCT
15	51	GGCATTGGCG	TCGGGCATTN	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CCTGTTCAAA	TTCATCATCG	GCGTACTCAA
	201	TANCCCCGAA	AAGATGCAGC	GTTTCGGTTC	GGCGCGTAAA	GGCCGCAAGG
	251	CCGCGCTTGC	TTGAACAAG	GCGGGTTTGG	CGTATTTTGA	AGGGCGTTTT
20	301	GAAAGGCGG	AACTTGAAGC	CTCGCGCGTA	TTGGGAAACA	AAGAGGCGGG
	351	GGATAACCGG	ACTTTGGCAT	TGATGTTGGG	CGCACATGCC	GCCGGGCAGA
	401	TGGAAAACAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCAACTG
	451	CCGGAAGAGC	AGCAGCTTTC	CCGTTATCTT	TTGTTGGCGG	AATCGGCGTT
	501	GAACCGGCGC	GATTACGAAG	CGGCGGAAGC	CAATCTTCAT	GCGGCGGCGA
25	551	AGATGAATGC	CAACCTTACG	CGCCTCGTGC	GTCTGCAACT	TCGTTACGCT
	601	TTCGACAGGG	GCGACGCGTT	GCAGGTTCTG	GCAAAAACCG	AAAAANTTTC
	651	CAAGGCGGGC	GCGTNGGGCA	AATCGGAAAT	GGAACGGTAT	CAAAATTGGG
	701	CATACCGCCG	CCAGCTGNCG	GATGCTGCCG	ATGCCGCGCG	TTTGAAAACC
	751	TGCCTGAAGC	GGATTCCCGA	CAGCCTCAAA	AACGGGGAAT	TGAGCGTATC
30	801	GGTTGCGGAA	AAGTACGAAC	GTTTGGGACT	GTATGCCGAT	GCGGTCAAAT
	851	GGGTCAAACA	GCATTATCCG	CACAACCGCC	GACCCGAACT	TTTGGAAGCN
	901	TTTGTCGAAA	GCGTGCCTT	TTTGGGCGAA	CGCGATCAGC	AGAAAGCCAT
	951	CGATTTTGCC	GATGCTTGGC	TGAAAGAACA	GCCCCGATAAT	GCGCTTCTGC
	1001	TGANGTATCT	CGGTGCGGCT	GCCTACGGCC	GCAAACTTTG	GGGCAAGGCA
35	1051	AAAGGCTACC	TTGAAGCGAG	CATTGCATTA	AAGCCGAGTA	TTTCCGCGCG
	1101	TTTGGTTCTG	GCAAAGGTTT	TTGACGAAAC	CGGAGAACCG	CAGAAGGCGG
	1151	AGGCGCAGCG	CAACTTGGTT	TTGGCAAGCG	TTGCCGAGGA	AAACCGNCCT
	1201	TCCGCCGAAA	CCCATTGA			

This encodes a protein having amino acid sequence <SEQ ID 754>:

40	1	MKTVVWIVVL	FAAAXGLALA	SGIXTGDVYI	VLGQTMLRIN	LHAFVLGSLI
	51	AVVVWYFLFK	FIIGVLNXPE	KMQRFGSARK	GRKAALALNK	AGLAYFEGRF
	101	EKAELASRV	LGNKEAGDNR	TLALMLGAHA	AGQMENIELR	DRYLAEIAKL
	151	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH	AAAKMNANLT	RLVRLQLRYA
	201	FDRGDALQVL	AKTEKXSKAG	AXGKSEMERY	QNWAYRRQLX	DAADAAALKT
45	251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWWKQHYP	HNRRPELLEA
	301	FVESVRFLGE	RDQQAIDFA	DAWLKEQPDN	ALLLXYLGRL	AYGRKLWGKA
	351	KGYLEASIAL	KPSISARLVL	AKVDFETGEP	QKAEAQRNLV	LASVAEENRP
	401	SAETH*				

ORF100a and ORF100-1 show 95.1% identity in 406 aa overlap:

50		10	20	30	40	50	60
	orf100a.pep	MKTVVWIVVLFAAAXGLALASGIXTGDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK					
	orf100-1	MKTVVWIVVLFAAAVGLALASGIYTGDDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK					
55		10	20	30	40	50	60
	orf100a.pep	FIIGVLNXPEKMQRFGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLGNKEAGDNR					
	orf100-1	FIIGVLNIPEKMQRFGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLVNKEAGDNR					
60		70	80	90	100	110	120
	orf100a.pep	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAEEANLH					
	orf100-1	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAEEANLH					
65		130	140	150	160	170	180
	orf100a.pep	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAEEANLH					
	orf100-1	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAEEANLH					

-422-

		130	140	150	160	170	180
5	orf100a.pep	190	200	210	220	230	240
	orf100-1	190	200	210	220	230	240
10	orf100a.pep	250	260	270	280	290	300
	orf100-1	250	260	270	280	290	300
15	orf100a.pep	310	320	330	340	350	360
	orf100-1	310	320	330	340	350	360
20	orf100a.pep	370	380	390	400		
	orf100-1	370	380	390	400		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF100 shows 93.3% identity over a 386 aa overlap with a predicted ORF (ORF100ng) from *N.gonorrhoeae*:

30	orf100.pep	MKTVVWIVVLF	AAA	AVGLALASGIY	TG	DVYIVLGQ	TMLRINLHAF	VLGSLIAV	VVWYFLFK	60					
	orf100ng	MKTVVWIVVLF	AAA	AVGLALASGIY	TG	DVYIVLGQ	TMLRINLHAF	VLGSLIAV	VVWYFLFK	60					
35	orf100.pep	FIIGVLN	IP	EKMQRF	GSARKG	XXXXLALN	KAGLAY	FEGRFE	KAELEAS	RVLNK	VGRDNR	120			
	orf100ng	FIIGVLN	IP	ENMR	RRSGS	ARKGRKA	ALALN	KAGLAY	FEGRFE	KAELEAS	RVLN	KNEAGDNR	120		
40	orf100.pep	TLALMLX	AHAAGQ	MENIX	RD	RYLAEI	AKLPEK	QQLS	RYLLA	ESALN	RRDYE	AAE	ANLH	180	
	orf100ng	TLALMLG	AHAAGQ	MENIE	LRD	RYLAEI	AKLPEK	QQLS	RYLLA	ESALN	RRDYE	AAE	ANLH	180	
45	orf100.pep	AAAKM	NANL	TRLV	RLXIR	YAFDR	GDALQ	VLAKTE	KL	SKAGAL	GKSE	MERYQ	NWAY	RRQLA	240
	orf100ng	AAAKM	NANL	TRLV	RLQ	LYAFDR	GDALQ	VLAKTE	KL	SKAGAL	GKSE	MERYQ	NWAY	RRQMA	240
50	orf100.pep	DAADAA	AALK	TCLK	RIPD	SLKNG	ELSV	SAEKY	ERLGL	YADAV	KWVKQ	HYP	PNRR	PELLEA	300
	orf100ng	DAADAA	AALK	TCLK	RIPD	SLKNG	ELSV	SAEKY	ERLGL	YADAV	KWVKQ	HYP	PHNRR	PELLEA	300
55	orf100.pep	FVESV	RFLG	EREQ	QKAID	FADAW	LKEQ	PDNALL	MYLGR	LAFGR	KLWG	KAKG	YLEAS	IAL	360
	orf100ng	FVESV	RFLG	EREQ	QKAID	FADSW	LKEQ	PDNALL	MYLGR	LAYGR	KLWG	KAKG	YLEAS	IAL	360
55	orf100.pep	KPSIS	ARL	VLT	KVFE	IGEP	QKAE	AH							386
	orf100ng	KPSI	PAR	L	VLA	KVFE	TAQ	SQKAE	AQRN	L	VLA	S	VAG	ENR	PSAETR

The complete length ORF100ng nucleotide sequence <SEQ ID 755> is:

	1	ATGAAACGG	TAGTCTGGAT	TGTTGCTCTG	TTTGCCGCCG	CCGTCGGACT
60	51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCCA	CGTGATATC	GTACTCGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CCTGTTTAA	TTCATCATCG	GCGTACTCAA
	201	TATCCCCGAA	AATATCGCGC	GTCCGGTTC	GGCGCGGAAA	GGCCGCAAGG
	251	CCGCGCTTGC	CTTGAATAAG	GCGGGTTTGG	CGTATTTCGA	AGGGCGTTT
	301	GAAAAGGCGG	AACTCGAAGC	CTCTCGAGTG	TTGGGCAACA	AAGAGGCCGG
65	351	AGACAACCGG	ACTTTGGCAT	TGATGCTGGG	CGCGCACGCG	GCAGGACAGA
	401	TGGAATAAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCAACTG

-423-

451 CCGGAAAAAC AGCAGCTTTC CCGCTATCTT CTGCTGGCGG AATCGGCGTT  
 501 AAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA  
 551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCC  
 601 TTCGATCGGG GCGATGCGTT GCAGTTCTG GCAAAAaccG AAAAAGCTTT  
 651 CAAGGCGGGC GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATTGGG  
 701 CATAACGCCG CCAGATGGCG GATGCTGCCG ATGCCGCCGC TTTGAAACC  
 751 TGCCTGAAGC GGATTCCCGA CAGCCTCAA ACGGGGAAT TGagcGTATC  
 801 GGTTGCGGAA AAGTACGAAC GTTTGGGACT GTATGCCGAT GCGGTCAAAT  
 851 GGGTCAAACA GCATTATCCG CACAACCGCC GCCCGAGCT TTTGGAAGCC  
 901 TTTGTCGAAA GCGTGCGCTT TTTGGGCGAG CGCGAACAGC AGAAAGCCAT  
 951 CGATTTTGCC GATTCTTGGC TGAAGAACA GCCCGATAAC GCGCTTCTGC  
 1001 TGATGTATCT CGGCCGGCTC GCCTACGGCC GCAAACTTTG GGGTAAGGCA  
 1051 AAAGGCTACC TTGAAGCGAG TATTGCACTG AAGCCGAGTA TTCCGGCGCG  
 1101 TTTGGTGTG GCAAAGGTTT TTGACGAAAC CGCACAGTCG CAAAAGCCG  
 1151 AAGCACAGCG CAACTTGCTT TTGGCAAGCG TTGCCGGGGA AAACCGCCCT  
 1201 TCCGCCGAAA CCCGTTGA

This encodes a protein having amino acid sequence <SEQ ID 756>:

1 MKTVVWIVVL FAAVGLALA SGIYTGdVYI VLQQTMLRIN LHAFVLGSLI  
 51 AVVVWYFLFK FIIGVLNIPE NMRRSGSARK GRKAALALNK AGLAYFEGRF  
 101 EKAELEASRV LGNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAEIAKL  
 151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAAMNANLT RLVRQLRYA  
 201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQMA DAADAAALKT  
 251 CLKRIPDSLK NGELSVSAE KYERLGLYAD AVKWKQHYH HNRPELLEA  
 301 FVESVRFLGE REQQKIDFA DSWLKEQPDN ALLMYLGR LAYGRKLWGKA  
 351 KGYLEASIAL KPSIPARLVL AKVFDETAQS QKAEQARNLV LASVAGENRP  
 401 SAETR\*

ORF100ng and ORF100-1 show 95.3% identity in 402 aa overlap:

		10	20	30	40	50	60
30	orf100-1.pep	MKTVVWIVVLFAAAVGLALASGIYTGdVYI	VLQQTMLRINLHAFVLGSLIAVVVWYFLFK				
	orf100ng	MKTVVWIVVLFAAAVGLALASGIYTGdVYI	VLQQTMLRINLHAFVLGSLIAVVVWYFLFK				
		10	20	30	40	50	60
35	orf100-1.pep	FIIGVLNIPEKMQRFSGSARKGRKAALALNKAGLAYFEGRFEKAELEASRV	LGNKEAGDNR				
	orf100ng	FIIGVLNIPENMRRSGSARKGRKAALALNKAGLAYFEGRFEKAELEASRV	LGNKEAGDNR				
		70	80	90	100	110	120
40	orf100-1.pep	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAAEANLH					
	orf100ng	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAAEANLH					
		130	140	150	160	170	180
45	orf100-1.pep	AAAMNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA					
	orf100ng	AAAMNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQMA					
		190	200	210	220	230	240
50	orf100-1.pep	DAADAAALKTCLKRIPDSLKNGELSVSAEKYERLGLYADAVKWKQHYHNRPELLEA					
	orf100ng	DAADAAALKTCLKRIPDSLKNGELSVSAEKYERLGLYADAVKWKQHYHNRPELLEA					
		250	260	270	280	290	300
55	orf100-1.pep	FVESVRFLGEREQQKIDFADAWLKEQPDNALLMYLGRLAYGRKLWGKAKGYLEASIAL					
	orf100ng	FVESVRFLGEREQQKIDFADSWLKEQPDNALLMYLGRLAYGRKLWGKAKGYLEASIAL					
		310	320	330	340	350	360
60	orf100-1.pep	KPSISARLVLAKVFDEIGEPQKAEQARNLVLEAVSDDERHAALQHSX					
	orf100n	KPSIPARLVLAKVFDETAQSQKAEQARNLVLASVAGENRPSAETRX					
		370	380	390	400		



370 380 390 400

Based on this analysis, including the presence of a putative leader sequence, a putative transmembrane domain, and a RGD motif, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 90

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 757>

```

10      1  ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCTGTG
      51  GTTTGCAGGG CTGTTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
     101  TTGATGTGCC GCGCGGCAAT CCCGAGTATG TGCGTCTGTC GGGCATGGCG
     151  GTGCGGCTGT ACCGTTTTAT GTCGCCGTTG GGCTTCGGCG CGGTCGTGTT
     201  CGGCGCGGCG ATACCGTTTG CCGCCGGCTG GTGGGGCAGC GGCTGGGTAC
     251  ACGTCAAAC TGTGTTGGGC TTGATGCTCT TGGCTTACCA GTTGATTGTC
     15  301  GCGGTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
     351  CTGGTACCGC GTGTTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
     401  TGTATsTGGT CGTGTTCAAA CCGTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 758; ORF102>:

```

20      1  MMFSWFKLFH LFFVISWFAG LFYLPRIEVDN MAMIDVPRGN PEYVRLSGMA
      51  VRLYRFMSPL GFGAVVFGAA IPFAAGWWGS GWVHVKLCLG LMLLAYQLYC
     101  GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYXVVFV PF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 759>:

```

25      1  ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCTGTG
      51  GTTTGCAGGG CTGTTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
     101  TTGATGTGCC GCGCGGCAAT CCCGAGTATG TGCGTCTGTC GGGCATGGCG
     151  GTGCGGCTGT ACCGTTTTAT GTCGCCGTTG GGCTTCGGCG CGGTCGTGTT
     201  CGGCGCGGCG ATACCGTTTG CCGCCGGCTG GTGGGGCAGC GGCTGGGTAC
     251  ACGTCAAAC TGTGTTGGGC TTGATGCTCT TGGCTTACCA GTTGATTGTC
     301  GCGGTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
     30  351  CTGGTACCGC GTGTTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
     401  TGTATCTGGT CGTGTTCAAA CCGTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 760; ORF102-1>:

```

35      1  MMFSWFKLFH LFFVISWFAG LFYLPRIEVDN MAMIDVPRGN PEYVRLSGMA
      51  VRLYRFMSPL GFGAVVFGAA IPFAAGWWGS GWVHVKLCLG LMLLAYQLYC
     101  GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYLVVFK PF*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with HP1484 hypothetical integral membrane protein of *H. pylori* (accession number AE000647)

ORF102 and HP1484 show 33% aa identity in 143aa overlap:

```

40      orf102  3  FSWFKLFHLFFVISWFAGLFYLPRIEVDNMAMIDVPRGNPEYVRLSGMAVRLYRFMSPLGF 62
      F W K FH+  VISW A LFYLPRI+EV A      +      V++      +LY F++
     HP1484  8  FLWVKAHFHVIIVISWMAALFYLPRLFYVHAENAHKKEFVGVVQIQEK--KLYSFIASPM 65

      orf102  63  GAVVFGAAIPFAAG---WWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWY 119
      G +      +      +      GW+H KL L ++LLAY YC +R +      + R+Y
     45      HP1484  66  GFTLITGIIMLLIEPTLFKSGGWLHAKLALVLLLAYHFYCKKCMRELEKDPTRRNARFY 125

      orf102  120 RVFNEIPXXXXXXXXXXXXXFKPF 142
      RVFNE P      KPF
     HP1484  126 RVFNEAPTILMILIVILVVVKPF 148

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF102 shows 99.3% identity over a 142aa overlap with an ORF (ORF102a) from strain A of *N. meningitidis*:

5		10	20	30	40	50	60
	orf102.pep	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVN	MAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL		
	orf102a	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVN	MAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL		
10		10	20	30	40	50	60
	orf102.pep	GFGAVVFGAAIPFAAGWWGSGVHV	KLCGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
	orf102a	GFGAVVFGAAIPFAAGWWGSGVHV	KLCGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
15		70	80	90	100	110	120
	orf102.pep	GFGAVVFGAAIPFAAGWWGSGVHV	KLCGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
	orf102a	GFGAVVFGAAIPFAAGWWGSGVHV	KLCGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
20		130	140				
	orf102.pep	VFNEIPVLLMVAALYLVVFKPFX					
	orf102a	VFNEIPVLLMVAALYLVVFKPFX					

The complete length ORF102a nucleotide sequence <SEQ ID 761> is:

	1	ATGATGTTTT	CTTGGTTCAA	GCTGTTTAC	TTGTTTTTTG	TCATTTCTGTG
	51	GTTTGCAGGG	CTGTTTTACC	TGCCGAGGAT	TTTCGTCAAT	ATGGCGATGA
25	101	TTGATGTGCC	GCGCGGCAAT	CCCGAGTATG	TGCGTCTGTC	GGGCATGGCG
	151	GTGCGGCTGT	ACCGTTTTAT	GTCGCCGTTG	GGCTTCGGCG	CGGTCGTGTT
	201	CGGCGCGGCG	ATACCGTTTG	CCGCCGGCTG	GTGGGGCAGC	GGCTGGGTAC
	251	ACGTCAAAC	GTGTTTGGGC	TTGATGCTCT	TGGCTTACCA	GTTGTATTGC
	301	GGCGTGCTGC	TGCGCCGTTT	TCAGGATTAC	AGCAATGCTT	TTTCACACCG
30	351	CTGGTACCGC	GTGTTCAACG	AAATCCCCGT	GCTGCTGATG	GTTGCCGCGC
	401	TGTATCTGGT	CGTGTTCAAA	CCGTTTTGA		

This encodes a protein having amino acid sequence <SEQ ID 762>:

	1	MMFSWFKLFH	LFFVISWFAG	LFYLPRI	FVN	MAMIDVPRGN	PEYVRLSGMA
	51	VRLYR	FMSPL	GFGAVVFGAA	IPFAAGWWGS	GWVHV	KLCGLMLLAYQLY
35	101	GVLLRRFQDY	SNAFSHRWYR	VFNEIPVLLM	VAALYLVVFK	PF*	

ORF102a and ORF102-1 show complete identity in 142 aa overlap:

		10	20	30	40	50	60
	orf102a.pep	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVN	MAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL		
40	orf102-1	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVN	MAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL		
		10	20	30	40	50	60
	orf102a.pep	GFGAVVFGAAIPFAAGWWGSGVHV	KLCGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
45	orf102-1	GFGAVVFGAAIPFAAGWWGSGVHV	KLCGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
		70	80	90	100	110	120
	orf102a.pep	GFGAVVFGAAIPFAAGWWGSGVHV	KLCGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
	orf102-1	GFGAVVFGAAIPFAAGWWGSGVHV	KLCGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
50		130	140				
	orf102a.pep	VFNEIPVLLMVAALYLVVFKPFX					
	orf102-1	VFNEIPVLLMVAALYLVVFKPFX					

55 Homology with a predicted ORF from *N.gonorrhoeae*

ORF102 shows 97.9% identity over a 142 aa overlap with a predicted ORF (ORF102ng) from *N. gonorrhoeae*:

-426-

```

    orf102.pep  MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    orf102ng    MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDAPRGNPEYVRLSGMAVRLYRFMSPL  60

5   orf102.pep  GFGAVVFGAAIPFAAGWGWGSGVHVHVKLCGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    orf102ng    GFGAVVFGAAIPFAAGRWGSGVHVHVKLCGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR  120

10  orf102.pep  VFNEIPVLLMVAALYXVVFVKPF  142
               ||||||||||||||||||
    orf102ng    VFNEIPVLLMVAALYL VVFVKPF  142

```

The complete length ORF102ng nucleotide sequence <SEQ ID 763> is:

```

1   ATGATGTTTT CTTGTTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCTGTG
15  51  GTTTGCAGGG CTGTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
    101 TTGATGCGCC GCGCGGCAAT CCCGAGTATG TCGCCTGTG GGGGATGGCG
    151 GTGCGGTTGT ACCGTTTTAT GTCGCCTTTG GGTTCGGCG CGGTCGTGTT
    201 CGGCGCGGCG ATACCGTTTG CCGCcgggcg GTGGGCGagc ggctggGTTT
    251 ACGTCAAACGT GTGTTGGGC TTGATGCTCT TGGCTTATCA GTTGATTGC
    301 GCGGTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
20  351 CTGGTACCGC GTGTTCAACg aAATCCCCGT GCTGCTGATG GTTGCCGCGC
    401 TGTATCTGGT CGTGTTCAAA CCGTTTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 764>:

```

1   MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDAPRGN PEYVRLSGMA
25  51  VRLYRFMSPL GFGAVVFGAA IPFAAGRWGS GWVHVHVKLCGL LMLLAYQLYC
    101 GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYL VVFK PF*

```

ORF102ng and ORF102-1 show 98.6% identity in 142 aa overlap:

```

               10      20      30      40      50      60
    orf102-1.pep MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
    orf102ng     MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDAPRGNPEYVRLSGMAVRLYRFMSPL
               10      20      30      40      50      60

               70      80      90      100     110     120
    orf102-1.pep GFGAVVFGAAIPFAAGWGWGSGVHVHVKLCGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR
    orf102ng     GFGAVVFGAAIPFAAGRWGSGVHVHVKLCGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR
               70      80      90      100     110     120

               130     140
    orf102-1.pep VFNEIPVLLMVAALYL VVFKPFX
    orf102ng     VFNEIPVLLMVAALYL VVFKPFX
               130     140

```

In addition, ORF102ng shows significant homology to a membrane protein from *H. pylori*:

```

45  gi|2314656 (AE000647) conserved hypothetical integral membrane protein
    [Helicobacter pylori] Length = 148
    Score = 79.2 bits (192), Expect = 1e-14
    Identities = 50/147 (34%), Positives = 68/147 (46%), Gaps = 13/147 (8%)

50  Query: 3   FSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDAPRGNPEYVRLSGMAVRLYRFMSPLGF 62
    F W K FH+ VISW A LFYLPRI FV A + V++ +LY F++
    Sbjct: 8   FLWVKAHFHVI AVISWMAALFYLPRI LFVYHAENAHKKEFVG VVQIQEK--KLYSFIASPAM 65

55  Query: 63  GAVVFGAAIP-----FAAGRWGSGVHVHVKLCGLMLLAYQLYCGVLLRRFQDYSNAFS 115
    G + + F +G GW+H KL L ++LLAY YC +R + +
    Sbjct: 66  GFTLITGILMLLIEPTLFKSG----GWLHAKLALVLLLAYHFYCKKCMRELEKDPTRRN 121

    Query: 116 HRWYRVFNEIPXXXXXXXXXXXXXFKPF 142
    R+YRVFNE P KPF
60  Sbjct: 122 ARFYRVFNEAPTILMILIVILVVVKPF 148

```

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 91

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 765>:

```

5      1  ATGGCAAAAA TGATGAAATG GCGGGCTGTT GCGGCGGTCG CGGCGGCAGC
      51  GGGTTGGGGC GGATGGTCTT AACTGAAGCC CGAGCCGCAC GTGCTTGATA
     101  TTACGGAAC  GGTCAGGCGC GGC // .....
    ///.  ATTTCTGTTA CGATTTTGTC CGAACCGGAT ACGCCGATTA AGGCGAAGCT
      51  CGACAGCGTC GACCCCGGGC TGACCACGAT GTCGTCGGGC GGTACAACA
    10   101  GCAGTACGGA TACGGCTTCC AATGCGGTCT ACTATTATGC CCGTTCGTTT
      151  GTGCCGAATC CGGACGGCAA ACTCGCCACG GGGATGACGA CGCAGAATAC
     201  GGTGAAATC  GACGGCGTGA AAAATGTGCT GATTATTCCG TCGCTGACCG
     251  TGAAAAATCG CGGCGGCAAG GCGTTGTGTC GCGTGTGGG  TCGGACGGC
     301  AAGGCGGCGG AACGCGAAAT CCGGACCGGT ATGAGAGACA GTATGAATAC
    15   351  CGAAGTAAAA AGCGGGTTGA AAGAGGGGGA CAAAGTGGTC ATCTCCGAAA
      401  TAACCGCCGC CGAGCAACAG GAAAGCGGCG AACGCGCCCT AGGCGGCCCC
     451  CCGCGCCGAT AA
  
```

This corresponds to the amino acid sequence <SEQ ID 766; ORF85>:

```

20      1  MAKMMKWAAY AAVAAAAVWG GWS.LKPEPH VLDITETVRR G.....
      51  .....
     101  .....
     151  .....
     201  ..... I SFTILSEPDT
     251  PIKAKLDSVD PGLTTMSSGG YNSSTDASN AVYYYARFV PNPDGKLATG
    25   301  MTTQNTVEID GVKNVLIIPS LTVKNRGGKA FVRVLGADGK AAEREIRTGM
     351  RDSMNTEVKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR*
  
```

Further work revealed the further partial nucleotide sequence <SEQ ID 767>:

```

30      1  .GTATCGGTCG GCGCGCAGGC ATCGGGGCAG ATTAAGATAC TTTATGTCAA
      51  ACTCGGGCAA CAGGTTAAAA AGGGCGATTT GATTGCGGAA ATCAATTCTGA
     101  CCTCGCAGAC CAATACGCTC AATACGGAAG AATCCAAGTT GGAAACGTAT
     151  CAGGCGAAGC TGGTGTCCGC ACAGATTGCA TTGGGCAGCG CGGAGAAGAA
     201  ATATAAGCGT CAGGCGGCGT TATGGAAGGA AAACGCGACT TCCAAAGAGG
     251  ATTTGGAAG  CGCGCAGGAT GCGTTTGCCG CCGCCAAAGC CAATGTTGCC
     301  GAGCTGAAGG CTTTAATCAG ACAGAGCAA  ATTTCCATCA ATACCGCCGA
    35   351  GTCGGAATTG GGCTACACGC GCATTACCGC AACGATGGAC GGCACGGTGG
      401  TGGCGATTCT CGTGAAGAG GGCAGACTG TGAACGCGGC GCAGTCTACG
     451  CCGACGATTG TCCAATTGGC GAATCTGGAT ATGATGTTGA ACAAATGCA
     501  GATTGCCGAG GGCGATATTA CCAAGGTGAA GGCGGGGCAG GATATTTCTG
     551  TTACGATTTT GTCCGAACCG GATACGCCGA TTAAGGCGAA GCTCGACAGC
    40   601  GTCGACCCCG GGCTGACCAC GATGTCGTCG GGCGGTTACA ACAGCAGTAC
      651  GGATACGGCT TCCAATGCGG TCTACTATTA TGCCCGTTTCG TTTGTGCCGA
     701  ATCCGACGCG CAAACTCGCC ACGGGGATGA CGACGCAGAA TACGGTTGAA
     751  ATCGACGGCG TGAAAAATGT GCTGATTATT CCGTCGCTGA CCGTGAAAAA
     801  TCGCGGCGGC AAGGCGTTTG TCGCGGTGTT GGGTGCAGAC GGCAAGGCGG
    45   851  CGGAACGCGA AATCCGACCG GGTATGAGAG ACAGTATGAA TACCGAAGTA
      901  AAAAGCGGGT TGAAAGAGGG GGACAAAGTG GTCATCTCCG AAATAACCGC
     951  CGCCGAGCAA CAGGAAAGCG GCGAACGCGC CTTAGGCGGC CCGCCGCGCC
    1001  GATAA
  
```

This corresponds to the amino acid sequence <SEQ ID 768; ORF85-1>:

```

50      1  .VSVGAQASQ IKILYVKLGQ QVKKGDLIAE INSTSQNTNL NTEKSKLETY
      51  QAKLVSAQIA LGSAEKYKQR QALWKENAT SKEDLESAQD AFAAAKANVA
     101  ELKALIRQSK ISINTAESEL GYTRITATMD GTVVAILVEE GQTVNAAQST
     151  PTIVQLANLD MMLNKMQIAE GDITKVKAGQ DISFTILSEP DTPIKAKLDS
     201  VDPGLTTMSS GGYNSSTDAT SNAVYYARS FVNPDPGKLA TGMTTQNTVE
    55   251  IDGVKNVLI  PSLTVKNRGG KAFVRVLGAD GKAAEREIRT GMRDSMNTEV
     301  KSLKEGDKV VISEITAAEQ QESGERALGG PRR*
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF85 shows 87.8% identity over a 41aa overlap and 99.3% identity over a 153aa overlap with an ORF (ORF85a) from strain A of *N. meningitidis*:

5	orf85.pep	MAKMMKWA	10	20	30	40	
	orf85a	MAKMMKWA	10	20	30	40	50 60
10	orf85.pep	.....			80	90	100
	orf85a	TIVQLANL	210	220	230	240	250 260
15	orf85.pep	110	120	130	140	150	160
	orf85a	270	280	290	300	310	320
20	orf85.pep	170	180	190	200	210	220
	orf85a	330	340	350	360	370	380
30	orf85.pep	230					
	orf85a	390					

The complete length ORF85a nucleotide sequence <SEQ ID 769> is:

35	1	ATGGCAAAAA	TGATGAAATG	GGCGGCTGTT	GCGGCGGTCG	CGGCGGCAGC
	51	GGTTTGGGGC	GGATGGTCTT	ATCTGAAGCC	CGAGCCGCAG	GCTGCTTATA
40	101	TTACGGAAAC	GGTCAGGCGC	GGCGACATCA	GCCGGACGGT	TTCTGCAACA
	151	GGGGAGATTT	CGCGTCCAA	CCTGGTATCG	GTCGGCGCGC	AGGCATCGGG
45	201	GCAGATTAAG	AAACTTTATG	TCAAACCTCG	GCAACAGGTT	AAAAAGGGCG
	251	ATTTGATTGC	GGAAATCAAT	TCGACCTCGC	AGACCAATAC	GCTCAATACG
50	301	GAAAAATCCA	AATTGGAAC	GTATCAGGCG	AAGCTGGTGT	CGGCACAGAT
	351	TGCATTGGGC	AGCGCGGAGA	AGAAATATAA	GCGTCAGGCG	GCGTTGTGGA
55	401	AGGATGATGC	GACCGCTAAA	GAAGATTGTT	AAAGCGCACA	GGATGCGCTT
	451	GCCGCCGCCA	AAGCCAATGT	TGCCGAGCTG	AAGGCTCTAA	TCAGACAGAG
60	501	CAAAATTTCC	ATCAATACCG	CCGAGTCGGA	ATTGGGCTAC	ACGCGCATTA
	551	CCGCAACGAT	GGACGGCAGC	GTGGTGGCGA	TTCTCGTGGA	AGAGGGCGAG
65	601	ACTGTGAACG	CGGCGCAGTC	TACGCCGACG	ATTGTCCAAT	TGGCGAATCT
	651	GGATATGATG	TTGAACAAAA	TGCAGATTGC	CGAGGGCGAT	ATTACCAAGG
70	701	TGAAGGCGGG	GCAGGATATT	TCGTTTACGA	TTTTGTCCGA	ACCGGATACG
	751	CCGATTAAGG	CGAAGCTCGA	CAGCGTCGAC	CCCGGGCTGA	CCACGATGTC
75	801	GTCGGGCGGC	TACAACAGCA	GTACGGATAC	GGCTTCCAAT	GCGGTCTACT
	851	ATTATGCCCG	TTCGTTTGTG	CCGAATCCGG	ACGGCAAAC	CGCCACGGGG
80	901	ATGACGACGC	AGAATACGGT	TGAAATCGAC	GGTGTGAAAA	ATGTGCTGAT
	951	TATTCCGTCG	CTGACCGTGA	AAAATCGCGG	CGGCAGGGCG	TTTGTGCGCG
85	1001	TGTTGGGTGC	AGACGGCAAG	GCGGCGGAAC	GCGAAATCCG	GACCGGTATG
	1051	AGAGACAGTA	TGAATACCGA	AGTAAAAAGC	GGGTTGAAAG	AGGGGGACAA
90	1101	AGTGGTCATC	TCCGAAATAA	CCGCCGCCGA	GCAGCAGGAA	AGCGGCGAAC
	1151	GCGCCCTAGG	CGGCCGCGCG	CGCCGATAA		

This encodes a protein having amino acid sequence <SEQ ID 770>:

60	1	MAKMMKWA	AAVAAAAVWG	GWSYLKPEPQ	AAYITETVRR	GDISRTVSAT
	51	GEISPSNLVS	VGAQASGQIK	KLYVKLGQV	KKGDLIAEIN	STSQTNTLNT
65	101	EKSKLETYQA	KLVSQAIALG	SAEKKYKRQA	ALWKDDATAK	EDLESAQDAL
	151	AAAKANVAEL	KALIRQSKIS	INTAESELGY	TRITATMDGT	VVAILVEEQ
70	201	TVNAAQSTPT	IVQLANLDM	LNKMQIAEGD	ITKVKAGQDI	SFTILSEPD
	251	PIKAKLDSVD	PGLTTMSSSG	YNSSTDASN	AVYYYARFV	PNPDGKLATG
75	301	MTTQNTVEID	GVKNVLIIPS	LTVKNRGGRA	FVRVLGADGK	AAEREIRTCM

351 RDSMNTTEVK SGLKEGDKVVI SEITAAEQQE SGERALGGPP RR\*

ORF85a and ORF85-1 show 98.2% identity in 334 aa overlap:

5	orf85a.pep	30	40	50	60	70	80
	orf85-1	PQAAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKGDLIAE					
					10	20	30
10	orf85a.pep	90	100	110	120	130	140
	orf85-1	INSTSQTNLTNTEKSKLETYQAKLVSAQIALGSAEKKYKRAALWKDDATAKEDLESAQD					
		40	50	60	70	80	90
15	orf85a.pep	150	160	170	180	190	200
	orf85-1	ALAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST					
		100	110	120	130	140	150
20	orf85a.pep	210	220	230	240	250	260
	orf85-1	PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS					
25		160	170	180	190	200	210
30	orf85a.pep	270	280	290	300	310	320
	orf85-1	GGYNSSTDASNAVYYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG					
		220	230	240	250	260	270
35	orf85a.pep	330	340	350	360	370	380
	orf85-1	RAFVRVLGADGKAAEREIRTGMRDSMNTTEVKSGLKEGDKVVI SEITAAEQQESGERALGG					
		280	290	300	310	320	330
40	orf85a.pep	390					
	orf85-1	PPRRX					
		PPRRX					

Figure 19D shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF85a..

#### Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF85 shows a high degree of identity with a predicted ORF (ORF85ng) from *N.gonorrhoeae*:

	ORF85	1	MAKMMKWAAVA AAAAAVWGWS.LKPEPHVLDITETVRRG.....	40
	ORF85ng	1	MAKMMKWAAVA AAAAAVWGWSY LKPEPQAA YITEAVRRGDISRTVSAT	50
50	ORF85		.....ISFTILSEPDT	250
	ORF85ng	201	TVNAAQSTPTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDT	250
55	ORF85	251	PIKAKLDSVDPGLTTMSSGGYNSSTDASNAVYYYARSFVNPDPGKLATG	300
	ORF85ng	251	PIKAKLDSVDPGLTTMSSGGYNSSTDASNAVYYYARSFVNPDPGKLATG	300
60	ORF85	301	MTTQNTVEIDGVKNVLIIPSLTVKNRGGKAFVRVLGADGKAAEREIRTGM	350
	ORF85ng	301	MTTQNTVEIDGVKNVLIIPSLTVKNRGGKAFVRVLGADGKAVEREIRTGM	350
	ORF85	152	RDSMNTTEVKSGLKEGDKVVI SEITAAEQQESGERALGGPPRR	393
65	ORF85ng	351	KDSMNTTEVKSGLKEGDKVVI SEITAAEQQESGERALGGPPRR	393

The complete length ORF85ng nucleotide sequence <SEQ ID 771> is:

```

1  ATGGCAAAAA TGATGAAATG GGCGGCTGTT GCGGCGGTCG CGGCGGCaac
51  GGTTTGGGGC GGATGGTCTT ATCTGAAGCC CGAACCGCAG GCTGCTTATA
101 TTACGGAAac ggTCAGGCGC GGCGATATCA GCCGGACGGT TTCCGCGACG
5  151  GgcgAGATT  CGCCGTCCAA CCTGGTATCG GTCGGCGCGC AGGCTTCGGG
201  GCAGATTAAA AAGCTTTATG TCAAACTCGG GCAACAGGTC AAAAAGGGCG
251  ATTTGATTGC GGAAATCAAT TCGACCACGC AGACCAACAC GATCGATATG
301  GAAAAATCCA AATTGGAAAC GTATCAGGCG AAGCTGGTGT CGGCACAGAT
351  TGCATTGGGC AGCGCGGAGA AGAAATATAA GCCTCAGGCG GCCTTGTGGA
10  401  AGGATGATGC GACCTCTAAA GAAGATTGG AAAGCGCGCA GGATGCGCTT
451  GCCGCCGCCA AAGCCAATGT TGCCGAGTTG AAGGCTTTAA TCAGACAGAG
501  CAAAATTTCC ATCAATACCG CCGAGTCGGA TTTGGGCTAC ACGCGCATTa
551  CCGCGACGAT GGACGGCACG GTGGTGGCGA TTCCCGTGGA AGAGGGGcAG
601  ACTGTGAACG CGGCGCAGTC TACGCCGACG ATTGTCCAAT TGGCGAATCT
15  651  GGATATGATG TTGAACAAAA TGCAGATTGC CGAGGGCGAT ATTACCAAGG
701  TGAAGGCGGG GCAGGATATT TCGTTTACGA TTTTGTCCGA ACCGGATACG
751  CCGATTAAGG CGAAGCTCGA CAGCGTCGAC CCCGGGCTGA CCACGATGTC
801  GTCGGGCGGC TACAACAGCA GTACGATAC GGCTTCCAAT GCGGTCTATT
851  ATTATGCCCG TTCGTTTGTG CCGAATCCGG ACGGCAAACT CGCCACGGGG
20  901  ATGACGACGC AGAATACGGT TGAATCGAC GGTGTGAAAA ATGTGTTGCT
951  TATTCCGTCG CTGACCGTGA AAAATCGCGG CGGCAAGGCG TTCGTACGCG
1001 TGTGTTGGTGC GGACGGCAAG GCAGTGAAC GCGAAATCCG GACCGGTATG
1051 AAAGACAGTA TGAATACCGA AGTGAAGAGC GGGTTGAAAG AGGGGGACAA
1101 AGTGGTCATC TCCGAAATAA CCGCCGCCGA GCAGCAGGAA AGCGGCGAAC
25  1151 GCGCCCTAGG CGGCCCGCCG CGCCGATAA

```

This encodes a protein having amino acid sequence <SEQ ID 772>:

```

1  MAKMMKWAav AAVAAAAVWG GWSYLKPEPQ AAYITEAVRR GDISRTVSAT
51  GEISPSNLVS VQAQASGQIK KLYVKLGQOV KKGDIAEIN STTQNTIDM
30  101  ESKLETYQA KLVSAQIALG SAEKKYKRQA ALWKDDATSK EDLESAQDAL
151  AAAKANVAEL KALIRQSKIS INTAESDLGY TRITATMDGT VVAIPVEEGQ
201  TVNAAQSTPT IVQLANLDM LNKMQIAEGD ITKVKAGQDI SFTILSEPDT
251  PIKAKLDSVD PGLTTMSSG YNSSTDASN AVYYARSFV PNPDGKLATG
301  MTTQNTVEID GVKNVLLIPS LTVKNRGGKA FVRVLGADGK AVEREIRTGM
351  KDSMNTVEKS GLKEGDKVVI SEITAEQQE SGERALGGPP RR*

```

ORF85ng and ORF85-1 show 96.1% identity in 334 aa overlap:

```

30      40      50      60      70      80
orf85ng  PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQOVKKGDIAE
40  orf85-1  |||||
          VSVGAQASGQIKILYVKLGQOVKKGDIAE
          10      20      30

90      100     110     120     130     140
orf85ng  INSTTQNTIDMEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKDDATSKEDLESAQD
45  orf85-1  ||||:||||:|||||
          INSTSQNTNLNTEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKENATSKEDLESAQD
          40      50      60      70      80      90

150     160     170     180     190     200
orf85ng  ALAAAKANVAELKALIRQSKISINTAESDLGYTRITATMDGTVVAIPVEEGQTVNAAQST
50  orf85-1  |:|||||
          AFAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST
          100     110     120     130     140     150

210     220     230     240     250     260
orf85ng  PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS
55  orf85-1  |||||
          PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS
          160     170     180     190     200     210

270     280     290     300     310     320
orf85ng  GGYNSSTDASNNAVYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG
60  orf85-1  |||||
          GGYNSSTDASNNAVYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG
          220     230     240     250     260     270

330     340     350     360     370     380

```

-431-

```

orf85ng      KAFVRVLGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVIS EITAAEQQESGERALGG
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf85-1      KAFVRVLGADGKAAEREIRTGMRDSMNTEVKSGLKEGDKVVIS EITAAEQQESGERALGG
                280      290      300      310      320      330

orf85ng      390
              PPRRX
              ||||
orf85-1      PPRRX

```

10 In addition, ORF85ng shows significant homology to an *E.coli* membrane fusion protein:

gi|1787104 (AE000189) o380; 27% identical (27 gaps) to 332 residues from  
membrane fusion protein precursor, MTRC\_NEIGO SW: P43505 (412 aa) [Escherichia  
coli] Length = 380

Score = 193 bits (485), Expect = 2e-48

Identities = 120/345 (34%), Positives = 182/345 (51%), Gaps = 13/345 (3%)

```

Query: 29  PQAA YITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKGDLIAE 88
          P   Y T  VR GD+ ++V ATG++      V VGAQ SGQ+K L V +G +VKK  L+
Sbjct: 41  PVPTYQTLIVRPGDLQQSVLATGKLDALRKVDVGAQVSGQLKTLSSAIGDKVKKDQLLGV 100

```

```

Query: 89  INSTTQTNTIDMEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKDDATSKEXXXXXXX 148
          I+   N I  ++ L  +A+  A+  L  A  Y RQ  L +  A S++
Sbjct: 101 IDPEQAENQIKEVEATLMELRAQRQQAELKLARVTYSRQRLAQTKA VSSQDLDTAAT 160

```

```

Query: 149 XXXXXXXXXXXXXXXXIRQSKISINTAESDLGYTRITATMDGTVVAIPVEEGQTVNAAQST 208
          I++++ S++TA+++L YTRI A M G V I  +GQTV AAQ
Sbjct: 161 EMAVKQAQIGTIDAQIKRNQASLDTAKTNLDYTRIVAPMAGEVTQITTLQGQTVIAAQA 220

```

```

Query: 209 PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS 268
          P I+ LA++ ML K Q++E D+  +K GQ  FT+L +P T  + ++ V P
Sbjct: 221 PNILTLADMSAMLVKAQVSEADVHLKPGQKAWFTVLGDPLTRYEGQIKDVLP----- 273

```

```

Query: 269 GGYNSSTD TASNAVYYAR SFVPNPDGK LATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG 328
          + +  ++A++YYAR VNP+G L  MT Q  +++ VKNVL IP  + + G
Sbjct: 274 -----TPEKVND AIFYYARFEVNPNGLLRLDMTAQVHIQLTDVKNVLTIPLSALGDPVG 328

```

```

Query: 329 KAFVRV-LGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVIS E 372
          +V L  +G+  ERE+  G ++  + E+  GL+ GD+VVI E
Sbjct: 329 DNRYKVLLRNGETREREVTIGARNDTDVEIVKGLEAGDEVVIGE 373

```

40 Based on this analysis, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*,  
and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF85-1 (40.4kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above.

The products of protein expression and purification were analyzed by SDS-PAGE. Figure 19A  
shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein

45 was used to immunise mice, whose sera were used for Western blot (Figure 19B), FACS analysis  
(Figure 19C), and ELISA (positive result). These experiments confirm that ORF85-1 is a  
surface-exposed protein, and that it is a useful immunogen.

### Example 92

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 773>:

```

50      1  ..ATTCCCGCCA CGATGACATT TGAACGCAGC GGCAATGCTT AAAAAATCGT
      51  TTCGACGATT AAAGTGCCGC TATACAATAT CCGTTTCGAG TCCGGCGGTA
     101  CGGTTGTCGG CAATACCCGT CACCCTACCT ACTATAGAGA CATACGCAGG
     151  GGCAAACGTG ATGCGGAAGc CAAATTCGCC GACgGcAGCG TAACTTACGG
     201  CAAAGCGGGC GAGAGCAAAA CCGAGCAAAG CCCCAGGCT ATGGATTTGT

```



-432-

5  
251 TCACGCTTGC CTGGCAGTTG GCGGCAAATG ACGCGAAACT CCCCCGGGG  
301 CTGAAAATCA CCAACGGCAA AAACTTTAT TCCGTCGGCG GTTTGAATAA  
351 GCGGGGTACA GGAAAATACA GCATAGGCGG CGTGGAACC GAAGTCGTCA  
401 AATATCGGGT GCGGCGGGC GACGATGCGG TAATGTATT cTTGCGACCG  
451 TCCCTGAACA ATATTCCGGC ACAAATCGGC TATACCGACG ACGGCAAAAC  
501 CTATACGCTG AAACCTCAAT CGGTGCAGAT CAACGGCCAG GCAGCCAAAC  
551 CGTAA

This corresponds to the amino acid sequence <SEQ ID 774; ORF120>:

10  
1 ..IPATMTFERS GNAYKIVSTI KVPLYNIRFE SGGTVVGNTL HPTYRDIRR  
51 GKLYAEAKFA DGSVTYKAG ESKTEQSPKA MDLFTLAWQL AANDAKLPPG  
101 LKITNGKKLY SVGGLNKAGT GKYSIGGVET EVVKYRVRRG DDAVMYFFAP  
151 SLNNIPAQIG YTDDGKTYTL KLKSVQINGQ AAKP\*

Further work revealed the complete nucleotide sequence <SEQ ID 775>:

15  
1 ATGATGAAGA CTTTTAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC  
51 CCTGCCGTGC GCGTATGCGG CAGGGCTGCC CCAATCCGCC GTGCTGCACT  
101 ATTCCGGCAG CTACGGCATT CCCGCCACGA TGACATTTGA ACGCAGCGGC  
151 AATGCTTACA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG  
201 TTTCGAGTCC GCGGTACGG TTGTCCGCAA TACCCTGCAC CCTACTACT  
251 ATAGAGACAT ACGCAGGGGC AAATGTATG CGGAAGCCAA ATTCGCCGAC  
301 GGCAGCGTAA CTTACGGCAA AGCGGGCGAG AGCAAACCG AGCAAAGCCC  
351 CAAGGCTATG GATTTGTTC CGCTTGCTG GCAGTTGGCG GCAAATGACG  
401 CGAAACTCCC CCCGGGGCTG AAAATCACCA ACGGCAAAA ACTTTATTCC  
451 GTCGGCGGTT TGAATAAGGC GGGTACAGGA AAATACAGCA TAGGCGCGT  
25  
501 GGAAACCGAA GTCGTCAAT ATCGGGTGC GCGCGGCGAC GATGCGGTAA  
551 TGTATTTCTT CGCACCCTCC CTGAACAATA TTCCGCGACA AATCGGCTAT  
601 ACCGACGACG GCAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA  
651 CGGCCAGGCA GCCAAACCGT AA

This corresponds to the amino acid sequence <SEQ ID 776; ORF120-1>:

30  
1 MMKTFKNIFS AAILSALPC AYAAGLPQSA VLHYSYSGYI PATMTFERSG  
51 NAYKIVSTIK VPLYNIRFES GGTVVGNTLH PTYYRDIRRG KLYAEAKFAD  
101 GSVTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS  
151 VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DAVMYFFAPS LNNIPAQIGY  
201 TDDGKTYTLK LKSVQINGQA AKP\*

Computer analysis of this amino acid sequence gave the following results:

### 35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF120 shows 92.4% identity over a 184aa overlap with an ORF (ORF120a) from strain A of *N. meningitidis*:

40  
orf120.pep  
orf120a  
10 20 30 40 50 60  
IPATMTFERSGNAYKIVSTIKVPLYNIRFE  
|||| : || |||||  
45  
orf120.pep SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYKAGESKTEQSPKAMDFTLAWQL  
orf120a SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYKAXXXXXXQSPKAMDFTLAWQL  
70 80 90 100 110 120  
50  
orf120.pep AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE EVVKYRVRRGDDAVMYFFAP  
orf120a AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE EVVKYRVRRGDDAVMYFFAP  
130 140 150 160 170 180  
55

-433-

```

                    160      170      180
orf120.pep  SLNNIPAQIGYTDGKTYTLKLKSVQINGQAAKPX
            |||
orf120a     SLNNIPAQIGYTDGKTYTLKLKSVQINGQAAKPX
5           190      200      210      220

```

The complete length ORF120a nucleotide sequence <SEQ ID 777> is:

```

1  ATGATGAAGA CTTTTAAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC
51 CCTGCCGTGC GCGTATGCGG CAGGGCTGCC CNAATCCGCC GTGCTGCACT
101 ATTCCGGCAG CTACGGCATT CCCGCCACNA NNANNTNNGN ACNNGNGNC
151 AATGCTTNCA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
201 TTCGAGTCC GCGCGTACGG TTGTCGGCAA TACCCTGCAC CCTACCTACT
251 ATAGAGACAT ACGCAGGGGC AAATGTATG CGGAAGCCAA ATTCGCCGAC
301 GGCAGCGTAA CCTACGGCAA AGCGGNNNNN ANCNNNNNNG NGCAAAGCCC
15 351 CAAGGCTATG GATTTGTTC ACGCTGCGTG GCAGTTGGCG GCAAATGACG
401 CGAAACTCCC CCCGGGGCTG AAAATCACCA ACGGCAAAAA ACTTTATTCC
451 GTCGGCGGTT TGAATAAGGC GGGTACAGGA AAATACAGCA TAGGCGGCGT
501 GGAAACCGAA GTCGTCAAAT ATCGGGTGCG GCGCGGCGAC GATGCGGTAA
551 TGTATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
20 601 ACCGACGACG GCAAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA
651 CGGCCAGGCA GCCAAACCGT AA

```

This encodes a protein having amino acid sequence <SEQ ID 778>:

```

1  MMKTFKNIFS AAILSAAALPC AYAAGLPXSA VLHYSYGI PATXXXXXXX
51 NAXKIVSTIK VPLYNIRFES GGTVVGNLTH PTYYRDIRRG KLYAEAKFAD
25 101 GSVTYGKAXX XXXXQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
151 VGGLNKAGTG KYSIGGVETE VVKYRVRGD DAVMYFFAPS LNNIPAQIGY
201 TDDGKTYTLK LKSVQINGQA AKP*

```

ORF120a and ORF120-1 show 93.3% identity in 223 aa overlap:

```

                    10      20      30      40      50      60
30 orf120a.pep  MMKTFKNIFS AAILSAAALPCAYAAGLPXSAVLHYSYGI PATXXXXXXXXNAXKIVSTIK
orf120-1       MMKTFKNIFS AAILSAAALPCAYAAGLPQSAVLHYSYGI PATMTFERSGNAYKIVSTIK
                    10      20      30      40      50      60

35 orf120a.pep  VPLYNIRFESGGTVVGNLTHPTYYRDIRRGKLYAEAKFADGSVTYGKAXXXXXXQSPKAM
orf120-1       VPLYNIRFESGGTVVGNLTHPTYYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAM
                    70      80      90      100     110     120

40 orf120a.pep  DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRGD
orf120-1       DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRGD
                    130     140     150     160     170     180

45 orf120a.pep  DAVMYFFAPSLNNIPAQIGYTDGKTYTLKLKSVQINGQAAKPX
orf120-1       DAVMYFFAPSLNNIPAQIGYTDGKTYTLKLKSVQINGQAAKPX
                    190     200     210     220

50 orf120a.pep  DAVMYFFAPSLNNIPAQIGYTDGKTYTLKLKSVQINGQAAKPX
orf120-1       DAVMYFFAPSLNNIPAQIGYTDGKTYTLKLKSVQINGQAAKPX
                    190     200     210     220

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF120 shows 97.8% identity over 184 aa overlap with a predicted ORF (ORF120ng) from *N.gonorrhoeae*:

```

55 orf120.pep  IPATMTFERSGNAYKIVSTIKVPLYNIRFE 30
orf120ng     SAAILSAALPCAYAARLPQSAVLHYSYGI PATMTFERSGNAYKIVSTIKVPLYNIRFE 69

60 orf120.pep  SGGTVVGNLTHPTYYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAMD LFTLAWQL 90
orf120ng     SGGTVVGNLTHPAYKDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAMD LFTLAWQL 129

```

```

orf120.pep  AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETEVLVKYRVRRGDDAVMYFFAP 150
|||||
orf120ng    AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETEVLVKYRVRRGDDTVTYFFAP 189
5
orf120.pep  SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKP 184
|||||
orf120ng    SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKP 223

```

The complete length ORF120ng nucleotide sequence <SEQ ID 779> is:

```

10      1  ATGATGAAGA CTTTAAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC
      51  CCTGCCGTGC GCGTATGCGG CAAGGCTACC CCAATCCGCC GTGCTGCACT
     101  ATTCGGCAG CTACGGCATT CCCGCCACGA TGACATTTGA ACGCAGCGGC
     151  AATGCTTACA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
     201  TTTCGAATCC GCGGGTACGG TTGTGCGCAA TACCCTGCAC CCTGCCTACT
15      251  ATAAAGACAT ACGCAGGGGC AAATGTATG CGGAAGCCAA ATTCGCCGAC
     301  GGCAGCGTAA CCTACGGCAA AGCGGGCGAG AGCAAAACCG AGCAAAGCCC
     351  CAAGGCTATG GATTGTGTTCA CGCTTGCTG GCAGTTGGCG GCAAATGACG
     401  CGAAACTCCC CCCGGTCTG AAAATCACCA ACGGCAAAA ACTTTATTCC
     451  GTCGGCGGCC TGAATAAGGC GGGTACGGGA AAATACAGCA TaggCGGCGT
20      501  GGAAACCGAA GTCGTCAAAT ATCGGGTGC GCGCGGCGAC GATACGGTAA
     551  CGTATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
     601  ACCGACGACG GCAAACCTA TACGCTGAAG CTCAAATCGG TGCAGATCAA
     651  CGGACAGGCC GCCAAACCGT AA

```

This encodes a protein having amino acid sequence <SEQ ID 780>:

```

25      1  MMKTFKNIFS AAILSAALPC AYAARLPQSA VLHYSGSYGI PATMTFERSG
     51  NAYKIVSTIK VPLYNIRFES GGTVVGNTLH PAYYKDIRRG KLYAEAKFAD
     101  GSVTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
     151  VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DTVTYFFAPS LNNIPAQIGY
     201  TDDGKTYTLK LKSVQINGQA AKP*

```

30 In comparison with ORF120-1, ORF120ng shows 97.8% identity in 223 aa overlap:

```

      10      20      30      40      50      60
orf120-1.pep MMKTFKNIFS AAILSAALPCAYAAGLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIK
|||||
35      10      20      30      40      50      60
orf120ng     MMKTFKNIFS AAILSAALPCAYAARLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIK
      70      80      90      100     110     120
orf120-1.pep VPLYNIRFESGGTVVGNTLHPYTYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAM
|||||
40      70      80      90      100     110     120
orf120ng     VPLYNIRFESGGTVVGNTLHPAYYKDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAM
      130     140     150     160     170     180
orf120-1.pep DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETEVLVKYRVRRGD
|||||
45      130     140     150     160     170     180
orf120ng     DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETEVLVKYRVRRGD
      190     200     210     220
orf120-1.pep DAVMYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX
|||
50      190     200     210     220
orf120ng     DTVTYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX

```

This analysis, including the presence of a putative leader sequence in the gonococcal protein suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 93

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 781>:

-435-

1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CCGTGGATGG GTGCCGGTGC  
 51 .GCGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA  
 101 CTCGGTTTGC GGTTCGGCGG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC  
 151 GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGCTGT  
 5 201 GATGGTGTTC TCCTTGATTT TGTTGTTGGC ATTATTGTTG ATTATCGTCC  
 251 CTATGCTGGT CGGGCAGTTC AACAAATTTGG CATCGCGCCT GCCCAATTA  
 301 ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG  
 351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC  
 10 401 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG  
 451 AGGCAGGGCG GCAATATT..

This corresponds to the amino acid sequence <SEQ ID 782; ORF121>:

1 MYRRKGRGIK PWMGAGXAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV  
 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF>NNLASRLPQL  
 101 IGFMTNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM  
 15 151 RQGGNI..

Further work revealed the complete nucleotide sequence <SEQ ID 783>:

1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CCGTGGATGG GTGCCGGTGC  
 51 GCGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA  
 101 CTCGGTTTGC GGTTCGGCGG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC  
 20 151 GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGCTGT  
 201 GATGGTGTTC TCCTTGATTT TGTTGTTGGC ATTATTGTTG ATTATCGTCC  
 251 CTATGCTGGT CGGGCAGTTC AACAAATTTGG CATCGCGCCT GCCCAATTA  
 301 ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG  
 351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC  
 25 401 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG  
 451 AGGCAGGGCG GCAATATTGT CAGCAGTATC GGCAACCTGC TGCTGCTTCC  
 501 CTTGCTGCTT TACTATTTCC TGCTGGATTG GCAGCGGTGG TCGTGGCGCA  
 551 TTGCCAAACT GGTTCGAGG CGTTTTGCCG GTGCTTATAC GCGCATTACA  
 601 GGCAATTTGA ACGAGGTATT GGGCGAATTT TTGCGCGGGC AGCTTCTGGT  
 30 651 AATGCTGATT ATGGGCTTGG TTTACGGTTT GGGATTGGTG CTGGTCGGGC  
 701 TGGATTCGGG GTTTGCCATC GGTATGCTTG CCGGTATTTT GGTGTTTGTG  
 751 CTTATCTCG GGGCGTTTAC GGGATTGCTG CTTGCCACCG TCGCCGCCTT  
 801 GCTCCAGTTC GGTTCGTGGA ACGGCATCCT ATCGGTTTGG GCGGTTTTTG  
 851 CCGTAGGACA GTTCTCGAA AGTTTTTCA TTACGCCGAA AATCGTGGGA  
 35 901 GACCGTATCG GGCTGTCGCC GTTTTGGGTT ATCTTTTCGC TGATGGCGTT  
 951 CGGGCAGCTG ATGGGCTTTG TCGGAATGTT GGCGGGATTG CCTTTGGCCG  
 1001 CCGTAACCTT GGTCTTGCTT CCGAGGGCG TGCAGAAATA TTTTGCCGGC  
 1051 AGTTTTTACC GGGCAGGTA G

This corresponds to the amino acid sequence <SEQ ID 784; ORF121-1>:

1 MYRRKGRGIK PWMGAGXAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV  
 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF>NNLASRLPQL  
 101 IGFMTNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM  
 151 RQGGNIVSSI GNLLLP LLL YFLLDWQRW SCGIKLVPR RFAGAYTRIT  
 201 GNLEVLGEF LRGQLVMI MGLVYGLGLV LVGLDSGFAI GMLAGILVFV  
 45 251 PYLGAFTGLL LATVAALLQF GSWNGILSVW AVFAVGQFLE SFFITPKIVG  
 301 DRIGLSPFWV IFSLMAFQQL MGFVGM LAGL PLAAVTLVLL REGVQKYFAG  
 351 SFYRGR\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF121 shows 98.7% identity over a 156aa overlap with an ORF (ORF121a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
55	orf121.pep	MYRRKGRGIK	PWMGAGXAFA	ALVWLVFALG	DTLTPFAVAA	VLAYVLDPLV	EWLQKKGLNR
	orf121a	MYRRKGRGIK	PWMDAGAAFA	ALVWLVFALG	DTLTPFAVAA	VLAYVLDPLV	EWLQKKGLNR
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf121.pep	ASASMSVMVF	SLILLALLL	IIVPMLVGQF	NNLASRLPQL	IGFMTNTLLP	WLKNTIGGYV

-436-

```

      |||
orf121a  ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
              70      80      90      100      110      120

5      130      140      150
orf121.pep EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNI
      |||
orf121a  EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
              130      140      150      160      170      180

10     orf121a  SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLD SGFAI
              190      200      210      220      230      240

```

The complete length ORF121a nucleotide sequence <SEQ ID 785> is:

```

15      1  ATGTATCGGA  GGAAAGGGCG  GGGCATCAAG  CCGTGGATGG  ATGCCGGTGC
      51  GGCGTTTGCC  GCCTTGGTCT  GGCTGGTTT  CGCGCTCGGC  GATACTTTGA
      101 CTCCGTTTGC  GGTTCGCGCG  GTGCTGGCGT  ATGTATTGGA  CCCTTTGGTC
      151 GAATGGTTGC  AGAAAAAGGG  TTTGAACCGT  GCATCCGCTT  CGATGTCTGT
      201 GATGGTGTTT  TCCTTGATT  TGTTGTTGGC  ATTATGTTG  ATTATGTCC
      251 CTATGCTGGT  CGGGCAGTTC  AACAAATTGG  CATCGCGCCT  GCCCCAATTA
      301 ATCGGTTTTA  TGCAGAACAC  GCTGCTGCCG  TGGTTGAAA  ATACAATCGG
      351 CGGATATGTG  GAAATCGATC  AGGCATCTAT  TATTGCGTGG  CTTCAGGCGC
      401 ATACGGGCGA  GTTGAGCAAC  GCGCTTAAGG  CGTGGTTTCC  CGTTTTGATG
      451 AGGCAGGGCG  GCAATATTGT  CAGCAGTATC  GGCAACCTGC  TGCTGCTTCC
      501 CTTGCTGCTT  TACTATTTCC  TGCTGGATTG  GCAGCGGTGG  TCGTGCGGCA
      25  551 TTGCCAACT  GGTTCGAGG  CGTTTTGCCG  GTGCTTATAC  GCGCATTACA
      601 GGCAATTTGA  ACGAGGTATT  GGGCGAATTT  TTGCGCGGGC  AGCTTCTGGT
      651 GATGCTGATT  ATGGGTTTGG  TTTACGGCTT  GGGGTGTTGG  CTGGTCGGGC
      701 TGGATTCGGG  GTTTGCAATC  GGTATGGTTG  CCGGTATTTT  GGTTTTGTT
      751 CCTATTTGG  GCGCGTTTAC  AGGACTGCTG  CTGGCAACCG  TCGCCGCCTT
      30  801 GCTCCAGTTC  GGTTCTGTGA  ACGGCATCTT  GGCTGTTTGG  GCGGTTTTTG
      851 CCGTAGGACA  GTTCTCGGAA  AGTTTTTCA  TTACGCCGAA  AATCGTGGGA
      901 GACCGTATCG  GCCTGTCGCC  GTTTTGGGTT  ATCTTTTCGC  TGATGGCGTT
      951 CGGGCAGCTG  ATGGGCTTG  TCGGAATGTT  GGCCGATTG  CCTTTGCCCG
      1001 CCGTAACCTT  GGTCTTGCTT  CGCGAGGGCG  TGCAGAAATA  TTTTGCCGGC
      35  1051 AGTTTTTACC  GGGGCAGGTA  G

```

This encodes a protein having amino acid sequence <SEQ ID 786>:

```

40      1  MYRRKGRGIK  PWM DAGAAFA  ALVWLVFALG  DTLTPFAVAA  VLAYVLDPLV
      51  EWLQKKGLNR  ASASMSVMVF  SLILLALLL  IIVPMLVGQF  NNLASRLPQL
      101  IGFMQNTLLP  WLKNTIGGYV  EIDQASIIAW  LQAHTGELSN  ALKAWFPVLM
      151  RQGGNIVSSI  GNLLLLPLLL  YYFLLDWQRW  SCGIAKLVPR  RFAGAYTRIT
      201  GNLNEVLGEF  LRGQLLVMLI  MGLVYGLGLV  LVGLD SGFAI  GMVAGILV FV
      251  PYLGAFTGLL  LATVAALLQF  GSWNGILAVW  AVFAVGQFLE  SFFITPKIVG
      301  DRIGLSPFWV  IFSLMAFGQL  MGFVGMLAGL  PLAAVTLVLL  REGVQKYFAG
      351  SFYRGR*

```

45 ORF121a and ORF121-1 show 99.2% identity in 356 aa overlap:

```

      10      20      30      40      50      60
orf121a.pep MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVA AVLAYVLDPLVEWLQKKGLNR
      |||
50  orf121-1  MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVA AVLAYVLDPLVEWLQKKGLNR
      10      20      30      40      50      60

      70      80      90      100      110      120
orf121a.pep ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
      |||
55  orf121-1  ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
      70      80      90      100      110      120

      130      140      150      160      170      180
orf121a.pep EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
      |||
60  orf121-1  EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
      130      140      150      160      170      180

      190      200      210      220      230      240
65  orf121a.pep SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLD SGFAI

```

-437-

	orfl21-1	SCGI AKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI	190	200	210	220	230	240
5	orfl21a.pep	GMVAGILVFVPYLGAFITGLLLATVAALLQFGSWNGILAVWAVFAVQGFLESFFITPKIVG	250	260	270	280	290	300
	orfl21-1	GMLAGILVFVPYLGAFITGLLLATVAALLQFGSWNGILSVWAVFAVQGFLESFFITPKIVG	250	260	270	280	290	300
10	orfl21a.pep	DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX	310	320	330	340	350	
15	orfl21-1	DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX	310	320	330	340	350	

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF121 shows 97.4% identity over a 156 aa overlap with a predicted ORF (ORF121ng) from *N.gonorrhoeae*:

20	orfl21.pep	MYRRKGRGIKPWMGAGXAFAALVWLVFALGDTLTPFAVAAYLAYVLDPLVEWLQKKGLNR	60
	orfl21ng	MYRRKGRGIKPWMGAGAAFAALVWLVYALGDTLTPFAVAAYLAYVLDPLVEWLQKKGLNR	60
25	orfl21.pep	ASASMSVMVFSLLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV	120
	orfl21ng	ASASMSVMVFSLLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV	120
	orfl21.pep	EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNI	156
30	orfl21ng	EIDQASIIAWFQAHTGELSNALKAWFPVLMKQGGNIVSTIGNLLPPLLLYYFLLDWHRW	180

An ORF121ng nucleotide sequence <SEQ ID 787> was predicted to encode a protein having amino acid sequence <SEQ ID 788>:

	1	MYRRKGRGIK	PWMGAGAAFA	ALVWLVYALG	DTLTPFAVAA	VLAYVLDPLV
35	51	EWLQKKGLNR	ASASMSVMVF	SLILLALLL	IIVPMLVGQF	NNLASRLPQL
	101	IGFMQNTLLP	WLKNTIGGYV	EIDQASIIAW	FQAHTGELSN	ALKAWFPVLM
	151	KQGGNIVSTI	GNLLLPPLL	YYFLLDWHRW	SCGIPKLVPR	RFAGAYTRIT
	201	GNLNKVGWKF	LRGQLGETE	RGAVVCRVGR	ECWEGGGARS	RPSDDGWPRW
	251	GGG*				

Further work revealed the following gonococcal DNA sequence <SEQ ID 789>:

40	1	ATGTATCGGA	GAAAAGGACG	GGGCATCAAG	CCGTGGATGG	GTGCCGGCGC
	51	GGCGTTTGCC	GCCTTGGTCT	GGCTGGTTTA	CGCGCTCGGC	GATACTTTGA
	101	CAATCGTTGC	GGTTGCGGCG	GTGCTGGCGT	ATGTGTTGGA	CCCTTTGGTC
	151	GAATGGTTGC	AGAAAAGGG	TTTGAACCGT	GCATCCGCTT	CGATGTCTGT
45	201	GATGGTGTTC	TCCTTGATTT	TGTTGTTGGC	ATTATTGTTG	ATTATTGTCC
	251	CTATGCTGGT	CGGGCAGTTC	AATAATTTGG	CATCTCGCCT	GCCCCAATTA
	301	ATCGGTTTTC	TGCAGAACAC	GCTGCTGCCG	TGGTTGAAAA	ATACAATCGG
	351	CGGATATGTG	GAAATCGATC	AGGCATCTAT	TATTGCGTGG	TTTCAGGCGC
	401	ATACGGGCGA	GTTGAGCAAC	GCGCTTAAGG	CGTGGTTTCC	CGTTTTGATG
50	451	AAACAGGGCG	GCAATATTGT	CAGCAGTATC	GGCAACCTGC	TGCTGCCGCC
	501	CTTGCTGCTT	TACTATTTC	TGCTGGATTG	GCAGCGGTGG	TCGTGCGGCA
	551	TCGCCAAACT	GGTCCGAGG	CGTTTTGCCG	GTGCTTATAC	GCGCATTACG
	601	GGTAATTGTA	ACGAGGTATT	GGGCGAATTT	TTGCGCGGTC	AGCTTCTGGT
	651	GATGCTGATT	ATGGGCTTGG	TTTACGGTTT	GGGATTGATG	CTAGTCGGAC
55	701	TGGATTCGGG	ATTTGCCATC	GGTATGGTTG	CCGGTATTTT	GGTGTTCGTC
	751	CCCTATTTGG	GTGCGTTTAC	GGGATTGCTG	CTTGCCACTG	TTGCAGCCTT
	801	GCTCCAGTTC	GGTTCGTGGA	ACGGAATCTT	GGCTGTTTGG	GCGGTTTTTG
	851	CCGTCGGTCA	GTTTCTCGAA	AGTTTTCCTA	TTACGCCGAA	AATTGTAGGA
	901	GACCGTATCG	GCCTGTCGCC	GTTTTGGGTT	ATCTTTTCGC	TGATGGCGTT
60	951	CGGAGAGCTG	ATGGGCTTTG	TCGGAATGTT	GGCCGGATTG	CCTTTGGCCG
	1001	CCGTAACCTT	GGTCTTCTT	CGCGAGGGCG	CGCAGAAATA	TTTGCCCGCG
	1051	AGTTTTTACC	GGGGCAGGTA	G		

This corresponds to the amino acid sequence <SEQ ID 790; ORF121ng-1>:

```

      1 MYRRKGRGIK PWMGAGAAFA ALVWLVAALG DTLTPFAVA VLAYVLDPLV
     51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL
    101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW FQAHTGELSN ALKAWFPVLM
5    151 KQGGNIVSSI GNLLLPPLL YYFLLDWQRW SCGIAKLVPR RFAGAYTRIT
    201 GNLNEVLGEF LRQQLLVMLI MGLVYGLGLM LVGLDSGFAI GMVAGILVVF
    251 PYLGAFTGLL LATVAALLQF GSWNGILAVW AVFAVGQFLE SFFITPKIVG
    301 DRIGLSPFWV IFSLMAFGEL MGFVGMLAGL PLAAVTLVLL REGAQKYFAG
    351 SFYRGR*

```

10 ORF121ng-1 and ORF121-1 show 97.5% identity in 356 aa overlap:

```

      10      20      30      40      50      60
or121-1.pep MYRRKGRGIKPWMGAGAAFAALVWLVFALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
15 or121ng-1 MYRRKGRGIKPWMGAGAAFAALVWLVAALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
      10      20      30      40      50      60

      70      80      90     100     110     120
or121-1.pep ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
20 or121ng-1 ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
      70      80      90     100     110     120

      130     140     150     160     170     180
or121-1.pep EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNIVSSIGNLLLPPLLYYFLLDWQRW
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
25 or121ng-1 EIDQASIIAWFQAHTGELSNALKAWFPVLMKQGGNIVSSIGNLLLPPLLYYFLLDWQRW
      130     140     150     160     170     180

      190     200     210     220     230     240
or121-1.pep SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
30 or121ng-1 SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLMLVGLDSGFAI
      190     200     210     220     230     240

      250     260     270     280     290     300
or121-1.pep GMLAGILVFVPYLGAFITGLLATVAALLQFGSWNGILSVWAVFAVGQFLESFFITPKIVG
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
35 or121ng-1 GMVAGILVFVPYLGAFITGLLATVAALLQFGSWNGILAVWAVFAVGQFLESFFITPKIVG
      250     260     270     280     290     300

      310     320     330     340     350
or121-1.pep DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
40 or121ng-1 DRIGLSPFWVIFSLMAFGELMGFVGMLAGLPLAAVTLVLLREGAQKYFAGSFYRGRX
      310     320     330     340     350

```

In addition, ORF121ng-1 shows homology to a permease from *H.influenzae*:

```

sp|P43969|PERM_HAEIN PUTATIVE PERMEASE PERM HOMOLOG Length = 349
Score = 69.9 bits (168), Expect = 2e-11
Identities = 67/317 (21%), Positives = 120/317 (37%), Gaps = 7/317 (2%)
50 Query: 26 VYALGDTLTPFAVAAVLAYVLDPLVEWL-QKKGLNRASASMSVMVFSXXXXXXXXXXXXXVP 84
      +Y GD + P +A VL+Y+L+ + +L Q R A++ + VP
Sbjct: 32 IYFFGDLIAPLLIALVLSYLLEIPINFLNQYLKCPRLATILIFGSFIGLAFFVLVLP 91

55 Query: 85 MLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYVE-IDQASIIAWFQAHTGELSNALK 143
      ML Q +L S LP + N WL N Y E ID + + + F + ++ +
Sbjct: 92 MLWNQTISLLSDLPAMF----NKSNEWLLNLPKNYPIDYSMVDSIFNSVREKILGFGE 147
Query: 144 AWFVPMKQGGNIVSSIGNXXXXXXXXXXXXXDWQRWSCGIAKLVPRRFAGAYTRITGNL 203

60      + + + N+VS D G+++ +P+ A+ R +
Sbjct: 148 SAVKLSLASIMNLVSLGIYAFLVPLMMFMFKDKSELLQGVSRLPKNRNLAFXRWK-EM 206

Query: 204 NEVLGEFLRGQXXXXXXXXXXXXXXXXXXXXXDSGFAIGMVAGILVFVPYXXXXXXXXXXXXX 263
      + + ++ G+ + + G+ V VPY
65 Sbjct: 207 QQQISNYIHGKLEILIVTLITYIIFLIFGLNYPLLLAFVGLSVLPYIGAVIVTIPVA 266

```

Query: 264 XXXXXQFGSWNGILAVWAVFAVGQFLESFFITPKIVGDRIGLSPFWVIFSLMAFGELMGF 323  
 QFG + FAV Q L+ + P + + + L P +I S++ FG L GF  
 Sbjet: 267 LVALFQFGISPTFWYIIIAFAVSQLLDGNLLVPYLFSEAVNLHPLIIIIISVLIFGGLWGF 326

5 Query: 324 VGMLAGLPLAAVTLVLL 340  
 G+ +PLA + ++  
 Sbjet: 327 WGVFFAIPLATLVKAVI 343

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the two proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 94

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 791>:

15 1 ..ACTGCTTTTT CGGCGGCGCT GCGCTTGAGT CCATCATGAC TCGTCATATT  
 51 TTTGTCCTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT  
 101 TTTGCACGTC CTGCCCCGCG CGTTCAAATG CGTACCAGCA ATACCGCCGC  
 151 CTGCGCCTCT ATGCCTTCCA TCCGCCGAG ATAGCCGAGT TTTTCGTTGG  
 201 TTTTGCTTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG  
 251 ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTGTGG GTTTCGTGTC  
 20 AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC  
 351 TTTGATACGC CGCACGCAA AGGACGCGGC TGTCCGCATC TTTGAACTCT  
 401 GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC  
 451 GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC  
 501 CGAGCAGCCC TTTTTCAAAT GGGATTTCAA CTCCGCCAAG TATCAG..

25 This corresponds to the amino acid sequence <SEQ ID 792; ORF122>:

1 ..TAFSAALRLS PSXLVIFLSF GKPYQQTAAI LTFECTSCPP RSNAYQQYRR  
 51 LRLYAFHPPE IAEFFVGFAF DVDARNVYAQ IGGDVGTHLR NVRRECGFLC  
 101 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGVGEM AADIAQTCT  
 151 EQRVGNVQQ RIGIGVSEQP FFKWDFNSAK YQ..

30 Further work revealed the complete nucleotide sequence <SEQ ID 793>:

1 ATATCGTACT GGGCAAGCAG TTCGCCGAT TTTTGGGAAG TAGATACCGC  
 51 GCCTTTGATT TTTTGCCGC TCTTACCCAA GGCTTCGATG AAAAAGTTGA  
 101 TGGTCGAGCC GGTACCGATG CCGATATATT CATTTCGGG TACGAATTCG  
 35 151 ACTGCTTTTT CGGCGGCGAT GCGCTTGAGT TCGTCTGTG TCGTCATATT  
 201 TTTGTCCTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT  
 251 TTTGCACGTC CTGCCCCGCG CGTTCAAATG CGTACCAGCA ATACCGCCGC  
 301 CTGCGCCTCT ATGCCTTCCA TCCGCCGAG ATAGCCGAGT TTTTCGTTGG  
 351 TTTTGCTTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG  
 401 ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTTTGG GTTTCGTGTC  
 40 AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC  
 501 TTTGATACGC CGCACGCAA AGGACGCGGC TGTCCGCATC TTTGAACTCT  
 551 GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC  
 601 GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC  
 651 CGAGCAGCCC TTTTTCAAAT GGGATTTCAA CTCCGCCAAG TATCAGCTTT  
 45 701 CTGCCTTCGG TCAGTTGGTG GACATCGTAG CCCTGTCCGA TACGGATGTT  
 751 CGTCATCGTT TGTGTTCTG A

This corresponds to the amino acid sequence <SEQ ID 794; ORF122-1>:

1 ISYWASSSPD FLEVDTAPLI FLPLLPKASM KKLMEVPVPM PIYSFSGTNS  
 50 51 TAFSAAMRLS SSCVVFIFLSF GKPYQQTAAI LTFECTSCPP RSNAYQQYRR  
 101 LRLYAFHPPE IAEFFVGFAF DVDARNVYAQ IGGDVGTHLR NVRREFGFLC  
 151 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGVGEM AADIAQTCT  
 201 EQRVGNVQQ RIGIGVSEQP FFKWDFNSAK YQLSAFGQLV DIVALSDDTV  
 251 RHR LCS\*

Computer analysis of this amino acid sequence gave the following results:



**Homology with a predicted ORF from *N.meningitidis* (strain A)**

ORF122 shows 94.0% identity over a 182aa overlap with an ORF (ORF122a) from strain A of *N*.

**meningitidis:**

[illegible]

The complete length ORF122a nucleotide sequence <SEQ ID 795> is:

	1	ATATCATATT	GGGCAAGCAG	TTCCTGGAT	TTTTTGAAG	TAGATACCGC
30	51	GCCTTTGATT	TTTTTGCCGC	TCTTACCCAA	GGCTTCGATG	AAAAAGTTGA
	101	TGGTCGAACC	GGTACCGATG	CCGATGTATT	CGTTTTCGGG	TACGAATTCTG
	151	ACTGCNTTTT	CGGCGGCGAT	GCGCTTGAGT	TCGTCTTGTG	TCGTCATATT
	201	TTTGTCTTT	GGGAAACCGT	ATCAACAAAC	AGCCGCCATC	TTAACATTTT
	251	TTNNAACGT	CTGCCCGCCG	CGTTCAAATC	CTTACCGACA	ATACCGCCGC
35	301	CTGCGACTCT	ATGCCTTCCA	TGCGCCCGAG	ATAACCGAGT	TTTTCGTTGG
	351	TTTTGCCTTT	GANGTTGACG	CACGAAATGT	CTATGCCCAA	ATCGCGCGCG
	401	ATGTTGGCAC	GCATTTGCCG	AATATGCGCG	GCGAGTTTGG	GTTTCTGTGC
	451	AATCAGCGTC	GTATCGACAT	TGACCGCCTG	CCAACCTTGC	GCCTGAACGC
	501	TTTGATACGC	CGCACGCAAA	AGGACGCGCG	TGTCGCGATC	TTTGAACCTCT
40	551	GCGGCGGTGT	CGGGGAAATG	GCTGCCGATA	TCGCCCCAAC	CTGCCGCACC
	601	GAGCAGCGCG	TCGGTAACGG	CGTGCAGCAG	CGCATCGGCA	TCGGAGTGTC
	651	CGAGCAGCCC	TTTTTCAAAT	GGGATTTCAA	CTCCGCCAAG	TATCAGCTTT
	701	CTGCCTTCGG	TCAGTTGGTG	GACATCGTAG	CCCTGTCCGA	TACGGATGTT
	751	CGTCATCGTT	TGTGTTCTGT	A		

This encodes a protein having amino acid sequence <SEQ ID 796>:

45	1	ISYWASSSLD	FLEVDTAPLI	FLPLLKASM	KKLMVEPVPM	PMYSFSGTNS
	51	TAFSAAMRLS	SSCVVIFLSF	GKPYQQTAAI	LTFFXTSCPP	RSNPYQQYRR
	101	LRLYAFHAPE	ITEFFVGFAF	XVDARNVYAQ	IGGDVGTHLR	NMRREFGFLC
	151	NHGRIDIDRL	PTLRLNALIR	RTQKDAAVRI	FELCGVGEM	AADIAQTCRT
50	201	EQRVGNVQQ	RIGIGVSEQP	FFKWDFNSAK	YQLSAFGQLV	DIVALSDDTV
	251	RHRLCS*				

ORF122a and ORF122-1 show 96.9% identity in 256 aa overlap:

```

55      10      20      30      40      50      60
orf122a.pep ISYWASSSLDFLEVDTAPLIFLPLLPKASMKKLMVEPVMPMYSFSGTNSTAFSAAMRLS
|||||||:|||||||
orf122-1 ISYWASSSPDFLEVDTAPLIFLPLLPKASMKKLMVEPVMPPIYSFSGTNSTAFSAAMRLS
      10      20      30      40      50      60

60      70      80      90     100     110     120
orf122a.pep SSCVVIFLSFGKPYQQTAAILTFXTSCPPRSNPYQQYRRLRLYAFHAPETEFFVGFAF
|||||||:|||||||
orf122-1 SSCVVIFLSFGKPYQQTAAILTFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAF

```

-441-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf122a.pep	XVDARNVYAQIGGDVGTHLRNMRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
	orf122-1	DVDARNVYAQIGGDVGTHLRNVRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
		130	140	150	160	170	180
10	orf122a.pep	FELCGGVGEMAADIAQTCRTEQRVGNGVQQRIGIGVSEQPFFKWFDFNSAKYQLSAFGQLV					
	orf122-1	FELCGGVGEMAADIAQTCRTEQRVGNGVQQRIGIGVSEQPFFKWFDFNSAKYQLSAFGQLV					
		190	200	210	220	230	240
15	orf122a.pep	DIVALSDTDVRHRLCSX					
	orf122-1	DIVALSDTDVRHRLCSX					
		250					
20							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF122 shows 89.6% identity over a 182 aa overlap with a predicted ORF (ORF122ng) from *N.gonorrhoeae*:

25	orf122.pep	TAFSAALRLSPSXLVIFLSFGKPYQQTAAI	30
	orf122ng	FLPLLPKASMKKLMVEPVPMPMYSFSGTNSTAFSAAMRLSSSCVIFLSFGKPYQQTAAI	80
	orf122.pep	LTFFCTSCPPRSNAYQQYRRLRLRYAFHPPEIAEFFVGFAFDVDARNVYAQIGGDVGTHLR	90
30	orf122ng	LTFFCTSWPPRSNPYQQYRRLRLRYAFHPPEIAEFFVGFAFDIDARNIDTQIGGDVGTHLR	140
	orf122.pep	NVRRECGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT	150
	orf122ng	NVRCEFGFLCNHGRIDIDHPLTLRLNALIRRTQKDAAVRIFELCGGVGKMAADVAQTCRT	200
35	orf122.pep	EQRVGNVQQRIGIGVSEQPFFKWFDFNSAKYQ	182
	orf122ng	EQRVGNVQQRVGRIMPEQPFFKWFDFNSAKYQLSAFGQLVDIVALSDTDIRHRLCS	256

The complete length ORF122ng nucleotide sequence <SEQ ID 797> is:

40	1	ATGTCGTACC	GGGCAAGCAG	TTCGCCGAT	TTTTTGGAGG	TTGAAACCGC
	51	GCCTTTGATT	TTTTTACCGC	TTTGCCCAA	GGCTTCGATG	AAGAAATTGa
	101	tgGTCGAACC	GgtACCGATG	CCGATGTATT	CGTTTTCGGG	TACGAATTCG
	151	ACTGCTTTTT	CGGCGGCGAT	GCGCttgAgt	TCgtcttgcg	TcgTCATATT
45	201	TTTAtccttt	gGGAaccct	atcaAcaAAc	agccgccatC	TTAACATTTT
	251	TTTGACGtc	ctggccgcg	cgttcaAATc	cgtaccaGca	ataccgccgc
	301	ctgcgctCT	AtgcCTTCCA	TCCGCCGAG	ATAGCCGAGT	TTTTCGTTGG
	351	TTTTGCCTTT	GATatTGACG	CACGAAATAT	CGatacCCAa	atcggcgcGCG
	401	ATGTTGGCAC	GCATTTGCGG	AATGTGCGGT	GCGAGTTTGG	GTTTCTGTGC
	451	AATCAGGTC	GTATCGACAT	TGACCACCTG	CCAACCCTGC	GCCTGAACGC
50	501	TTTGATACGC	CGCACGCAAA	AGGACGCGGC	TGTCCGCATC	TTTGAATCTT
	551	GCGGCGGTGT	CGGAAAATG	GCTGCCGATG	TCGCCCAAAC	CTGCCGCACC
	601	GAGCAGCg	tcggtaaCGG	CGTGACGAG	cgcgTcgGCA	TCCGAATGCC
	651	CGAGCAGCCC	TTTTTCAAAT	GGGATTTC	CTCCGCCAAG	TATCAGCTTT
	701	TGCCTTCGG	TCAATTGGTG	GACATCGTAG	CCCTGTCCGA	TACGGATATT
55	751	CGTCATCGTT	TGTGTTCTCG	A		

This encodes a protein having amino acid sequence <SEQ ID 798>:

	1	MSYRASSSPD	FLEVETAPLI	FLPLLPKASM	KKLMVEPVP	PMYSFSGTNS
	51	TAFSAAMRLS	SSCVIFLSF	GKPYQQTAAI	LTFFCTSWPP	RSNPYQQYRR
60	101	LRLYAFHPPE	IAEFFVGFAF	DIDARNIDTQ	IGGDVGTHLR	NVRCEFGFLC
	151	NHGRIDIDL	PTLRLNALIR	RTQKDAAVRI	FELCGGVGKM	AADVAQTCRT
	201	EQRVGNVQ	RVGIRMPEQP	FFKWFDFNSAK	YQLSAFGQLV	DIVALSDTD
	251	RHRLCS*				

ORF122ng and ORF122-1 show 92.6% identity in 256 aa overlap:

		10	20	30	40	50	60
	orf122-1.pep	ISYWASSSPDFLEVDTAPLIFLPLLPKASMKKLMVEPVPMPYISFSGTNSTAFSAAMRLS					
5	orf122ng	MSYRASSSPDFLEVETAPLIFLPLLPKASMKKLMVEPVPMPYISFSGTNSTAFSAAMRLS					
		10	20	30	40	50	60
	orf122-1.pep	SSCVVIFLSFGKPYQQTAAILTFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAF					
10	orf122ng	SSCVVIFLSFGKPYQQTAAILTFCTSWPPRSNPYQQYRRLRLYAFHPPEIAEFFVGFAF					
		70	80	90	100	110	120
	orf122-1.pep	DVDARNVYAQIGGDVGTHLRNVRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
15	orf122ng	DIDARNIDTQIGGDVGTHLRNVRCEFGFLCNHGRIDIDHPTLRLNALIRRTQKDAAVRI					
		130	140	150	160	170	180
	orf122-1.pep	FELCGGVGEMAAADIAQTCRTEQRVGNVQQRIGIGVSEQPFFKWDFNSAKYQLSAFGQLV					
20	orf122ng	FELCGGVGKMAADVAQTCRTEQRVGNVQQRVGIRMPEQPFFKWDFNSAKYQLSAFGQLV					
		190	200	210	220	230	240
25	orf122-1.pep	DIVALSDTDVRHRLCSX					
	orf122ng	DIVALSDTDIRHRLCSX					
30		250					

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 95

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 799>:

35	1	..GCCGGCGCGA	GTGCGAACAA	CATTTCGCG	CGTTTTCGCG	AAACACCCGT
	51	CGCTGTCAGC	GTTACCTGA	TCGGCACGGT	ACTTGCCGTC	ATGCTGCCCC
	101	TTACCGAATA	TGAAACTTC	CTGCTGCTTA	TCGGCTCGGT	ATTTGCGCCG
	151	ATGGGCGCGA	TTTGTATTGC	CGACTTTTTC	GTCTTGAAAC	GGCGTGA

This corresponds to the amino acid sequence <SEQ ID 800; ORF125>:

40	1	..AGASANNISA	RFAETPVAVS	VTLIGTVLAV	MLPVTEYENF	LLIGSVFAP
	51	MGGFDCRLFR	LETA*			

Further work revealed the complete nucleotide sequence <SEQ ID 801>:

	1	ATGTCGGGCA	ATGCCTCCTC	TCCTTCATCT	TCCTCCGCCA	TCGGGCTGAT
	51	TTGGTTCGGC	GCGGCGGTAT	CGATTGCCGA	AATCAGCACG	GGTACGCTGC
45	101	TTGCGCCTTT	GGGCTGGCAG	CGCGGTCTGG	CGGCTCTACT	TTTGGGTCAT
	151	GCCGTCGGCG	GCGCGCTGTT	TTTTGCGGCG	GCGTATATCG	GCGCACTGAC
	201	CGGACGCAGC	TCGATGGAAA	GCGTGCGCCT	GTCGTTCGGC	AAACGCGGTT
	251	CAGTGCTGTT	TTCCGTGGCG	AATATGCTGC	AACTGGCCGG	CTGGACGGCG
	301	GTGATGATT	ACGCCGGCGC	AACGGTCAGC	TCCGCTTTGG	GCAAAGTGTT
50	351	GTGGGACGGC	GAATCTTTTG	TCTGGTGGGC	ATTGGCAAAC	GGCGCGCTGA
	401	TTGTGCTGTG	GCTGGTTTTC	GGCGCACGCA	AAACAGGCGG	GCTGAAAACC
	451	GTTTCGATGC	TGCTGATGCT	GTTGGCGGTT	CTGTGGCTGA	GTGCCGAAGT
	501	CTTTTCCACG	GCAGGCAGCA	CCGCCGCACA	GGTTTCAGAC	GGCATGAGTT
	551	TCGGAACGGC	AGTCGAGCTG	TCCGCCGTGA	TGCCGCTTTC	CTGGCTGCCG
55	601	CTTGCCGCCG	ACTACACGCG	CCACGCGCGC	CGCCCGTTTG	CGGCAACCCT
	651	GACGGCAACG	CTCGCCTACA	CGCTGACCGG	CTGCTGGATG	TATGCCTTGG
	701	GTTTGGCAGC	GGCGTTGTTC	ACCGGAGAAA	CCGACGTGGC	AAAAATCCTG
	751	CTGGGCGCAG	GTTTGGGTGC	GGCAGGCATT	TTGGCGGTGC	TCCTCTCCAC

5, 

801	CGTTACCACA	ACGTTTCTCG	ATGCTTATTC	CGCCGCGCG	AGTGC GAACA
851	ACATTTCGCG	CGCTTTTGCG	GAAACACCG	TCGCTGTCG	CGTTACCCTG
901	ATCGGCACGG	TACTTGCCGT	CATGCTGCC	GTTACCGAA	ATGAAAACTT
951	CCTGCTGCTT	ATCGGCTCG	TATTTGCGC	GATGCGGCG	GTTTTGATTG
1001	CCGACTTTTT	CGTCTTGAA	CGCGTGAGG	AGATTGAAG	CTTTGACTTT
1051	GCCGGACTGG	TTCTGTGGCT	TGCGGGCTTC	ATCCTCTACC	GCTTCTGCT
1101	CTCGTCCGGC	TGGGAAAGCA	GCATCGGTCT	GACCGCCCC	GTAAATGTCTG
1151	CCGTTGCCAT	TGCCACCCTA	TCGGTACGCC	TTTTCTTTAA	AAAAACCCAA
1201	TCTTTACAAA	GGAAACCGTC	ATGA		

10 This corresponds to the amino acid sequence <SEQ ID 802; ORF125-1>:

15

1	<u>MSGNASSPSS</u>	<u>SSAIGLIWFG</u>	<u>AAVSIAEIST</u>	<u>GTLLAPLGWQ</u>	<u>RGLAALLLGH</u>
51	<u>AVGGALFFAA</u>	<u>AYIGALTGRS</u>	<u>SMESVRLSFG</u>	<u>KRGSVLFSVA</u>	<u>NMLQLAGWTA</u>
101	<u>VMIYAGATVS</u>	<u>SALGKVLWDG</u>	<u>ESFVWWALAN</u>	<u>GALIVLWLVF</u>	<u>GARKTGGLKT</u>
151	<u>VSMLLMLLAV</u>	<u>LWLSAEVFST</u>	<u>AGSTAAQVSD</u>	<u>YMSFGTAVEL</u>	<u>SAVMPLSWLP</u>
201	<u>LAADYTRHAR</u>	<u>RPFAATLTAT</u>	<u>LAYTLTGQWM</u>	<u>YALGLAAALF</u>	<u>TGETDVAKIL</u>
251	<u>LGAGLGAAGI</u>	<u>LAVVLSTVTT</u>	<u>TFLDAYSAGA</u>	<u>SANNISARFA</u>	<u>ETPVAVGVTL</u>
301	<u>IGTVLAVMLP</u>	<u>VTEYENFLLL</u>	<u>IGSVFAPMAA</u>	<u>VLIADFFVLK</u>	<u>RREEIEGFDF</u>
351	<u>AGLVLWLAFG</u>	<u>ILYRFLSSG</u>	<u>WESSIGLTAP</u>	<u>VMSAVAIATV</u>	<u>SVRLFEEKTO</u>
401	<u>SLORNPS*</u>				

20 Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF125 shows 76.5% identity over a 51aa overlap with an ORF (ORF125a) from strain A of *N. meningitidis*:

```

25      orf125.pep          AGASANNISARFAETPVAVSVTLIGTVLAV
                        ||:|||||:::||:||:::|||
    orf125a      KILLGAGLGAGILAVVLSTVTTFDAYSAGVSANNISAKLSEIPIAVAVAVVGTTLLAV
                250        260        270        280        290        300

30      orf125.pep          MLPVTEYENFLLLLIGSVVFAPMGGFDCRLRLETAX
                        :|||||||:
    orf125a      LLPVTEYENFLLLLIGSVFAPMAAVLIADFFVLKRREEIEG
                310        320        330        340

```

35 The ORF125a partial nucleotide sequence <SEQ ID 803> is:

	1	ATGTCGGGCA	ATGCCTCCTC	TCNTTCATCT	TCCGCCGCCA	TCCGGGTGAT
	51	TTGGTTCGGC	GCGGCGGTAT	CGATTGCCGA	AATCAGCACG	GGTACACTGC
	101	TTGCGCCTTT	GGGCTGGCAG	CGCGGTCTGG	CNGCTCTGCT	TTTGGGT CAT
40	151	GCCGTCGGCG	GCGCGCTGTT	TTTTGCGGCG	GCGTATATCG	GCGCACTGAC
	201	CGGACNCAN	TCGATGGAAA	GCGTGCGCCT	TCGTCTCGCG	AAACGCGGTT
	251	CAGTGCTGTT	TTCCGTGGCG	AATATGCTGC	AACTGGCGCG	CTGGACGGCG
	301	GTGATGATTT	ACGCCGGCGC	AACGGTCAGC	TCCGCTTTGG	GCAAAGTGTT
	351	GTEGGACGGC	GAATCTTTTG	TCTGGTGGGC	ATTGGCAAAC	GGCGCGCTGA
45	401	TTGTGCTGTG	GCTGGTTTTT	GGCGCACGCA	AAACAGGCGG	GCTGAAAACC
	451	GTTTCGATGC	TGCTGATGCT	GTTGGCGGTT	CTGTGGCTGA	GTCGCCGAAT
	501	NTTTTCCACG	GACGGCAGCA	CCGCCCGANN	GGTNNCAGTA	GGCATGAGTT
	551	TCGGAACGGC	AGTCGAGCTG	TCCGCCGTNA	TGCCGCTTTC	TTGGCTGCCG
	601	CTGGCCGCCG	ACTACACGCG	CCACGCGCGC	CGCCCGTTTG	CGGCAACCCT
50	651	GACGGCAACG	CTCGCCTACA	CGCTGACCGG	CTGCTGGATG	TATGCCTTTG
	701	GTTTGGCAGC	GGCGTTGTTT	ACCGGAGAAA	CCGACGTGGC	AAAAATCCTG
	751	CTGGGCGCAG	GTTTGGGTGC	GCGAGGCATT	TGTGGCGTCG	TCCTGTCGAC
	801	CGTTACCACC	ACTTTTCTCG	ATGCNTACTC	CGCCGGCGTA	AGTGCCAAAC
	851	ATATTTCCGC	CAAACTTTCG	GAAATACCNA	TCGCCGTTGC	CGTCGCCGTT
55	901	GTCGGCACAC	TGCTTGCCGT	CCTCCTGCCC	GTTACCGAAT	ATGAAAACTT
	951	CCTGCTGCTT	ATCGGTCTCG	TATTTGCGCC	GATGGCGCGG	GTTTTGATTG
	1001	CGCACTTTTT	CGCTTTGAAA	CGCGGTGAGG	AGATTGAARG	C..

This encodes a protein having the partial amino acid sequence <SEQ ID 804>:

1 MSGNASSXSS SAAIGLIWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH  
51 AVGGALFFAA AYIGALTGXX SMESVRLSFG KRGSVLFSSVA NMLQLAGWTA

-444-

101 VMIYAGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARKTGGLKT  
 151 VSMLLMLLAV LWLSAEXFST AGSTAAXVXD GMSFGTAVEL SAVMPLSWLP  
 201 LAADYTRHAR RPFATLTAT LAYTLTGCM YALGLAALF TGETDVAKIL  
 251 LGAGLGAAGI LAVVLSTVTT TFLDAYSAGV SANNISAKLS EIPAVAVAV  
 301 VGTLLAVLLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEG..

ORF125a and ORF125-1 show 94.5% identity in 347 aa overlap:

		10	20	30	40	50	60
10	orf125a.pep	MSGNASSXSSAAIGLIWFGAAVSIAEISTGTLAPLWQORGLAALLLGHAVGGALFFAA					
		:					
	orf125-1	MSGNASSPSSSSAIGLIWFGAAVSIAEISTGTLAPLWQORGLAALLLGHAVGGALFFAA					
		10	20	30	40	50	60
		70	80	90	100	110	120
15	orf125a.pep	AYIGALTGXSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG					
		:					
	orf125-1	AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG					
		70	80	90	100	110	120
20		130	140	150	160	170	180
	orf125a.pep	ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEXFSTAGSTAAXVXD					
		:					
	orf125-1	ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEVFSTAGSTAQVSD					
		130	140	150	160	170	180
25		190	200	210	220	230	240
	orf125a.pep	GMSFGTAVELSAVMPLSWPLAADYTRHARRPFAATLTATLAYTLTGCMYALGLAALF					
		:					
	orf125-1	GMSFGTAVELSAVMPLSWPLAADYTRHARRPFAATLTATLAYTLTGCMYALGLAALF					
30		190	200	210	220	230	240
		250	260	270	280	290	300
	orf125a.pep	TGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGVSANNISAKLSEIPIAVAVAV					
		:					
35	orf125-1	TGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGASANNISARFAETPVAVGVT					
		250	260	270	280	290	300
40		310	320	330	340		
	orf125a.pep	VGTLLAVLLPVTEYENFLLIGSVFAPMAAVLIADFFVLKRREEIEG					
		:      :					
	orf125-1	IGTVLAVMLPVTEYENFLLIGSVFAPMAAVLIADFFVLKRREEIEGDFDFAGLVLWLAGF					
		310	320	330	340	350	360

#### Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF125 shows 86.2% identity over a 65aa overlap with a predicted ORF (ORF125ng) from *N.gonorrhoeae*:

	orf125.pep	AGASANNISARFAETPVAVSVTLIGTVLAV	30
		:	
50	orf125ng	KILLGAGLGITGILAVVLSTVTTTFLDTYSAGASANNISARFAEIPVAVGVTLIRTVLAV	308
	orf125.pep	MLPVTEYENFLLIGSVFAPM-GGFDCLRFRLTA	64
		:	
	orf125ng	MLPVTEYKNFLLIRSVFGPMAGGFDCLRFCLKTA	343

55 An ORF125ng nucleotide sequence <SEQ ID 805> was predicted to encode a protein having amino acid sequence <SEQ ID 806>:

	1	MSGNASSPSS	SAAIGLVWFG	AAVSIAEIST	GTLAPLWQ	RGLAALLLGH
	51	AVGGALFFAA	AYIGALTGRS	SMESVRLSFG	KCGSVLFSVA	NMLQLAGWTA
	101	VMIYVGATVS	SALGKVLWDG	ESFVWWALAN	GALIVLWLVF	GARRTGGLKT
	151	VSMLLMLLAV	LWLSVEVFAS	SGTNAAPAVS	DGMTFGTAVE	LSAVMPLSWL
60	201	PLAADYTRQA	RRPFAATLTA	TLAYTLTGCM	MYALGLAAL	FTGETDVAKI
	251	LLGAGLGITG	ILAVVLSTVT	TTFLDTYSAG	ASANNISARF	AEIPVAVGVT
	301	LIRTVLAVML	PVTEYKNFLL	LIRSVFGPMA	GGFDCLRFCL	KTA*

Further work revealed the following gonococcal DNA sequence <SEQ ID 807>:

```

      1  ATGTCGGGCA ATGCCTCCTC TCCTTCATCT TCCGCCGCCA TCGGGCTGGT
    51  TTGGTTTCGGC GCGGCGGTAT CGATTGCCGA AATCAGCACG GGTACGCTGC
    101  TCGCCCCCTT GGGCTGGCAG CGCGGTCTGG CGGCCCTGCT TTTGGGTCAT
    151  GCGGTCGGCG GCGCGCTGTT TTTTGGCGCG GCGTATATCG GCGCACTGAC
    201  CGGACGCAGC TCGATGGAAA GTGTGCGCCT GTCGTTCCGC AAATGCGGTT
    251  CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGGCG
    301  GTGATGATTT ACGTCGGCGC AACGGTCAGC TCCGCTTGG GCAAAGTGTT
    351  GTGGGACGGC GAATCCTTTG TCTGGTGGGC ATTGGCAAAC GGCGCACTGA
    401  TCGTGCTGTG GCTGGTTTTT GGCGCACGCA GAACGGGCGG GCTGAAAACC
    451  GTTTCGATGC TGCTGATGCT GCTTGCCGTG TTGTGTTGA GCGTCGAAGT
    501  GTTCGCTTCG TCCGGCACAA ACGCCGCGCC CGCCGTTTCA GACGGCATGA
    551  CCTTCGGAAC GGCAGTCGAA CTGTCCGCCG TCATGCCGCT TTCCTGGCTG
    601  CCGCTGGCCG CCGACTACAC GCGCCAAAGC CGCCGCCCGT TTGCGGCAAC
    651  CCTGACGGCA ACGCTCGCCT ATACGCTGAC GGGCTGCTGG ATGTATGCCT
    701  TGGGTTTGGC GCGGCTCTG TTTACCGGAG AAACCGACGT GGCGAAAATC
    751  CTGTTGGGCG CGGGCTTGGG CATAACGGGC ATTCTGCGAG TCGTCCTCTC
    801  CACCGTTACC ACAACGTTTC TCGATACCTA TTCCGCCGGC GCGAGTCCGA
    851  ACAACATTTT CGCGCGTTTT GCGGAAATAC CCGTCGCTGT CGGCCTTACC
    901  CTGATCGGCA CCGTGCTTGC CGTCATGCTG CCCGTTACCG AATATAAAAA
    951  CTTCCTGCTG CTTATCGGCT CGGTATTTGC GCCGATGGCG GCGGTTTTGA
   1001  TTGCCGACTT TTTCTCTTAA AAACGGCGTG AGGAGATTGA AGGCTTTGAC
   1051  TTTGCCGGAC TGGTTCTGTG GCTGGCAGGC TTCATCCTCT ACCGCTTCCT
   1101  GCTCTCGTCC GGTGTTGGAAA GCAGCATCGG TCTGACCGCC CCCGTAATGT
   1151  CTGCCGTTGC CATTGCCACC GTATCGGTAC GCCTTTTCTT TAAAAAAACC
   1201  CAATCTTTAC AAAGGAACCC GTCATGA

```

This corresponds to the amino acid sequence <SEQ ID 808; ORF125ng-1>:

```

      1  MSGNASSPSS SAAIGLVWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
    51  AVGGALFFAA AYIGALTGRS SMESVRLSFG KCGSVLFSVA NMLQLAGWTA
    101  VMIYVGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARRTGGLKT
    151  VSMLLMLLAV LWLSVEVFAS SGTNAAPAVS DGMTFGTAVE LSAVMPLSWL
    201  PLAADYTRQA RRPFAATLTA TLAYTLTCGW MYALGLAAAL FTGETDVAKI
    251  LLGAGLGITG ILAVVLSTVT TTFLDTYSAG ASANNISARF AEIPVAVGVT
    301  LIGTVLAVML PVTEYKNFLL LIGSVFAPMA AVLIADFFVL KRREEIEGFD
    351  FAGLVWLWLAG FILYRFLSS GWESSIGLTA PVMSAVAIAT VSVRLFFKKT
    401  QSLQRNPS*

```

ORF125ng-1 and ORF125-1 show 95.1% identity in 408 aa overlap:

```

      10      20      30      40      50      60
    40  orf125-1.pep  MSGNASSPSSSSAIGLIWFGAAVVSIAEISTGTLLAPLGWQ RGLAALLLGHAVGGALFFAA
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
    orf125ng-1  MSGNASSPSSSSAIGLVWFGAAVVSIAEISTGTLLAPLGWQ RGLAALLLGHAVGGALFFAA
      10      20      30      40      50      60

      70      80      90      100     110     120
    45  orf125-1.pep  AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
    orf125ng-1  AYIGALTGRSSMESVRLSFGKCGSVLFSVANMLQLAGWTAVMIYVGATVSSALGKVLWDG
      70      80      90      100     110     120

      130     140     150     160     170     179
    50  orf125-1.pep  ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEVFSTAGSTAAQ-VS
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
    orf125ng-1  ESFVWWALANGALIVLWLVFGARRTGGLKTVSMLLMLLAVLWLSVEVFASSGTNAAPAVS
      130     140     150     160     170     180

      180     190     200     210     220     230     239
    55  orf125-1.pep  DGMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTG CWMYALGLAAAL
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
    orf125ng-1  DGMTFGTAVELSAVMPLSWLPLAADYTRQARRPFAATLTATLAYTLTG CWMYALGLAAAL
      190     200     210     220     230     240

      240     250     260     270     280     290     299
    60  orf125-1.pep  FTGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGASANNISARFAETPVAVGVT
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
    orf125ng-1  FTGETDVAKILLGAGLGITGILAVVLSTVTTTFLDTYSAGASANNISARFAEIPVAVGVT
      250     260     270     280     290     300

```

-446-

		300	310	320	330	340	350	359
5	orf125-1.pep	LIGTVLAVMLPVTEYENFLL	LIGSVFAPMAAVLIADFFVL	KRREEIEGDFAGLV	LWLAG			
	orf125ng-1	LIGTVLAVMLPVTEYKNFLL	LIGSVFAPMAAVLIADFFVL	KRREEIEGDFAGLV	LWLAG			
		310	320	330	340	350	360	
10	orf125-1.pep	FILYRFLSSGWESSIGLTAPVMSAVAIATVSVRLFFKKTQSLQRNPSX						
	orf125ng-1	FILYRFLSSGWESSIGLTAPVMSAVAIATVSVRLFFKKTQSLQRNPSX						
		370	380	390	400			

Based on this analysis, including the presence of putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 96

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 809>:

20	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCGGGAA	GGCTGACCGC
	51	GTTGCAGCTT	GCAGAACAAAG	GTTATCAGAT	TGCACTTTTC	GATAAAAGCT
	101	GCCGCCGGGG	CGAACACGCC	GCCGCCTATG	TAGCCGCCGC	CATGCTCGCG
	151	CCTGCAGCGG	A.ACGGTCGA	AGCCACGCCC	GAAGTGGTCA	GGCTGGGCAG
	201	GCAGAGCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCGTCTG	AACACGCACA
25	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TATGGCACGG	GCAGGACAAG
	301	CCATTATCCA	GCGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGT.AC GGA
	351	TGACGAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACCGC
	401	AACTCGGCGG	ACGTTTTTAA	GACGGCATCT	ACCTGCCGAC	CGAAGC.CAG
	451	CTCGACGGGC	GGCAATTATA	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCCGAA	GCCTGCAAG..

30 This corresponds to the amino acid sequence <SEQ ID 810; ORF126>:

1	MTRIAILGGG	LSGRLTALQL	AEQGYQIALF	DKSCRRGEHA	AAYVAAAMLA
51	PAAXTVEATP	EVVRLGRQSI	PLWRGIRCL	NTHMTMQENG	SLIVWHGQDK
101	PLSSEFVRHL	KRGXTDDEI	VRWRADDIAE	REPQLGGRFX	DGIYLPTEXQ
151	LDGRQLXSAL	ADALDELNVP	CHWEHECVPE	ACK...	

35 Further work revealed the complete nucleotide sequence <SEQ ID 811>:

	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCGGGAA	GGCTGACCGC
	51	GTTGCAGCTT	GCAGAACAAAG	GTTATCAGAT	TGCACTTTTC	GATAAAAGCT
	101	GCCGCCGGGG	CGAACACGCC	GCCGCCTATG	TGCGCGCCGC	CATGCTCGCG
	151	CCTGCAGCGG	AAGCGTCTGA	AGCCACGCCC	GAAGTGGTCA	GGCTGGGCAG
40	201	GCAGAGCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCGTCTG	AACACGCACA
	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TGTGGCACGG	GCAGGACAAG
	301	CCATTATCCA	GCGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGTAGCGGA
	351	TGACGAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACCGC
	401	AACTCGGCGG	ACGTTTTTCA	GACGGCATCT	ACCTGCCGAC	CGAAGGCCAG
45	451	CTCGACGGGC	GGCAATATT	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCCGAA	GGCCTGCAAG
	551	CCCAATACGA	CTGGCTGATC	GACTGCCGCG	GCTACGGCGC	AAAAACCGCG
	601	TGGAACCAAT	CCCCGAGCA	CACCAGCACC	CTGCGCGGCA	TACGCGGCGA
	651	AGTGGCGCGG	GTTTACACAC	CCGAAATCAC	GCTCAACCGC	CCCGTGCGTC
50	701	TGCTCCATCC	GCGTTATCCG	CTTACATCG	CCCCGAAAGA	AAACCACGTC
	751	TTCTGTCATCG	GCGCGACCCA	AATCGAAAGC	GAAAGCCAAG	CCCCCGCCAG
	801	CGTGGCTTCA	GGGTTGGAAC	TCTGTCCGCG	ACTCTATGCC	ATCCACCCCG
	851	CCTTCGGCGA	AGCCGACATC	CTCGAAATCG	CCACCGGCCT	GCGCCCCACG
	901	CTCAACCACC	ACAACCCCGA	AATCCGTTAC	AACCGCGCCC	GACGCCTGAT
55	951	TGAAATCAAC	GGCCTTTTCC	GCCACGGTTT	CATGATCTCC	CCCGCGGTAA
	1001	CCGCCGCCGC	CGCCAGATTG	GCACTGGCAC	TGTTTGACGG	AAAAGACGCG
	1051	CCCGAACGCG	ATAAGAAAG	CGGTTTGCGG	TATATCCGAA	GACAAGATTA
	1101	A				

This corresponds to the amino acid sequence <SEQ ID 812; ORF126-1>:

```

1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAMLA
51  PAEEAVEATP EVVRLGRQSI PLWRGIRCL NTHTMMQENG SLIVWHGQDK
101 PLSSEFVRHL KRGGVADDEI VRWRADDIAE REPQLGGRFS DGIYLPTEGQ
151 LDGRQILSAL ADALDELNVP CHWEHECVPE GLQAQYDWLI DCRGYGAKTA
201 WNQSPHTST LRGIRGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENHV
251 FVIGATQIES ESQAPASVRS GLELLSALYA IHPAFGEADI LEIATGLRPT
301 LNHHNPEIRY NRARRLIEIN GLFRHGFMIS PAVTAAARL AVALFDGKDA
351 PERDKESGLA YIRRD*

```

10 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF126 shows 90.0% identity over a 180aa overlap with an ORF (ORF126a) from strain A of *N. meningitidis*:

```

15  orf126.pep  10      20      30      40      50      60
      MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAYVAAAMLAPAAXTVEATP
      orf126a   10      20      30      40      50      60
      MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAYVAAAMLAPAAEVEATP

20  orf126.pep  70      80      90      100     110     120
      EVVRLGRQSIPLWRGIRCLNTHTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGXTDDEI
      orf126a   70      80      90      100     110     120
      EVVRLGRQXIPLWRGIRCHLKT PAMMXENGSLIVWHGQDKPLSNEFVRHLKRGGVADDXI

25  orf126.pep  130     140     150     160     170     180
      VRWRADDIAEREPQLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECVPE
      orf126a   130     140     150     160     170     180
      VRWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPE
30

```

The complete length ORF126a nucleotide sequence <SEQ ID 813> is:

```

1  ATGACCCGTA TCGCCATCCT CGGCGGCGGC CTCTCNGGAA GGCTGACCGC
51  ACTGCAGCTT GCAGAACAAAG GTTATCAGAT TGCACTTTTC GATAAAGGCT
101 GCCGCCGGGG CGAACACGCC GCCGCCTATG TTGCCGCGC CATGCTCGCG
151 CCTGCGGCGG AAGCGGTCTGA AGCCACGCCT GAAGTGGTCA GGCTGGGCAG
201 GCAGANCATC CCGCTTTGGC GCGGCATCCG ATGCCATCTG AAAACGCCTG
251 CCATGATGCA NGAAAACGCC AGCCTGATTG TGTGGCACGG GCAGGACAAA
301 CCTTTATCCA ACGAGTTCGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA
351 TGACNAAATC GTCGTTGGC GCGCCGACGA CATCGCCGAA CGCGAACCGC
401 AACTCGGCGG ACGTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG
451 CTCGACGGGC GGCAAATATT GTCTGCACTT GCCGACGCTT TGGACGAACT
501 GAACGTCCCC TGCCATTGGG AACACGAATG TGCCCCCGAA GACTTGCAAG
551 CCCAATACGA CTGGCTGATC GACTGCCGCG GCTACGGCGC AAAAACC GCG
601 TGAACCAAT CCCCCGANN NACCAGCACC CTGCGCGGCA TACGCGGCGA
45  AGTGGCGCGG GTTTACACAC CCGAAATCAC GCTCAACCGC CCCGTGCGCC
701 TGCTACACCC GCGCTATCCG CTNTACATCG CCCCAGAAAG AAACNCGTC
751 TCGTTCATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CACCTGCCAG
801 CGTGCGTTCC GGGCTGGAAC TCTTATCCGC ACTCTATGCC GTCCACCCCG
851 CCTTCGGCGA AGCCGACATC CTCGAAATCG CCACCGGCTT GCGCCCCACG
50  CTCAATCACC ACAACCCCGA AATCCGTAC AACCGCGCCC GACGCCTGAT
951 TGAAATCAAC GGCCTTTTCC GCCACGTTT CATGATCTCC CCCGCCGTAA
1001 CCGCCGCGCG CGTCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGANGCG
1051 CCCGAACGCG ATGAAGAAAG CGGTTTGGCG TATATCCGAA GACAAGATTA
1101 A

```

55 This encodes a protein having amino acid sequence <SEQ ID 814>:

```

1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAMLA
51  PAEEAVEATP EVVRLGRQXI PLWRGIRCHL KTPAMMXENG SLIVWHGQDK
101 PLSNEFVRHL KRGGVADDEI VRWRADDIAE REPQLGGRFS DGIYLPTEGQ
151 LDGRQILSAL ADALDELNVP CHWEHECAPE DLQAQYDWLI DCRGYGAKTA
60  201 WNQSPXST LRGIRGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENXV

```



-448-

251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIATGLRPT  
 301 LNHHNPEIRY NRARRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKXA  
 351 PERDEESGLA YIRRQD\*

ORF126a and ORF126-1 show 95.4% identity in 366 aa overlap:

5		10	20	30	40	50	60
	orf126a.pep	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP					
	orf126-1	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP					
10		10	20	30	40	50	60
	orf126a.pep	EVVRLGRQXIPLWRGIRCHLKTTPAMMXENGSLIVWHGQDKPLSNEFVRHLKRGGVADDDXI					
	orf126-1	EVVRLGRQSIPLWRGIRCLNTHMTMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI					
15		70	80	90	100	110	120
	orf126a.pep	VRWRADDIAEREPLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPE					
	orf126-1	VRWRADDIAEREPLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECVPE					
20		130	140	150	160	170	180
	orf126a.pep	DLQAQYDWLIDCRGYGAKTAWNQSPXXSTSLRGIRGEVARVYTPEITLNRPVRLHPRYP					
	orf126-1	GLQAQYDWLIDCRGYGAKTAWNQSPEHTSTSLRGIRGEVARVYTPEITLNRPVRLHPRYP					
25		190	200	210	220	230	240
	orf126a.pep	LYIAPKENXVFVIGATQIESESQAPASVRSGLLELLSALYAVHPAFGEADILEIATGLRPT					
	orf126-1	LYIAPKENHVFVIGATQIESESQAPASVRSGLLELLSALYAIHPAFGEADILEIATGLRPT					
30		250	260	270	280	290	300
	orf126a.pep	LNHHNPEIRYNRARRLIEINGLFRHGFMSPAVTAAAVRLAVALFDGKXAPERDEESGLA					
	orf126-1	LNHHNPEIRYNRARRLIEINGLFRHGFMSPAVTAAARLAVALFDGKDAPERDKESGLA					
35		310	320	330	340	350	360
	orf126a.pep	YIRRQDX					
	orf126-1	YIRRQDX					
40		310	320	330	340	350	360
	orf126a.pep	YIRRQDX					
	orf126-1	YIRRQDX					
45							

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF126 shows 90% identity over a 180 aa overlap with a predicted ORF (ORF126ng) from *N.gonorrhoeae*:

50	orf126.pep	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAXTVEATP	60
	orf126ng	MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTROGEHAAAYVAAAMLAPAAEAVEATP	60
	orf126.pep	EVVRLGRQSIPLWRGIRCLNTHMTMQENGSLIVWHGQDKPLSSEFVRHLKRGGXTDDEI	120
55	orf126ng	EVIRLGRQSIPLWRGIRCLNTHMTMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI	120
	orf126.pep	VRWRADDIAEREPLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECVPE	180
	orf126ng	VRWRADEIAEREPLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQ	180

60 An ORF126ng nucleotide sequence <SEQ ID 815> was predicted to encode a protein having amino acid sequence <SEQ ID 816>:

1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTROGEHA AAYVAAAMLA

-449-

51 PAEEAVEATP EVIRLGRQSI PLWRGIRCRL NTLTMMQENG SLIVWHGQDK  
 101 PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ  
 151 LDGRQILSAL ADALDELNVP CHWEHECAPQ DLQAQYDWVI DCRGYGAKTA  
 201 WNQSPEHTST LRGIRGEVRG FTRPKSRSTA PCACCTRAIR STSPRKKTTTS  
 251 SSSARPKSKA KAKPPPAYVP GWNSYPRSMPT STPPSAKPTS SKWRPGLRPT  
 301 LNHHNPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA  
 351 PERDEESGLA YIGRQD\*

Further work revealed the following gonococcal DNA sequence <SEQ ID 817>:

10 1 ATGACCCGTA TCGCCGTCCT CGGAGGCGGC CTTTCCGGAA GGCTGACCGC  
 51 ATTGCAGCTT GCAGAACAAAG GTTATCAGAT TGAACCTTTC GACAAGGGCA  
 101 CCCGCCAAGG CGAACACGCC GCCGCCTATG TTGCCGCCGC GATGCTCGCG  
 151 CCTGCGGCGG AAGCGGTCTGA GGCAACGCCC GAAGTCATCA GGCTGGGCAG  
 201 GCAGAGCATT CCGCTTTGGC GCGGCATCCG ATGCCGTCTG AACACGCTCA  
 15 251 CGATGATGCA GGAAAACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAG  
 301 CCATTATCCA GCGAGTTCGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA  
 351 TGACGAAATC GTCCGTTGGC GCGCCGATGA AATCGCCGAA CGCGAACCGC  
 401 AACTCGGCGG ACGTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG  
 451 CTCGACGGGC GGCAAAATATT GTCTGCACTT GCCGACGCTT TGGACGAACT  
 20 501 GAACGTCCCT TGCATTGGG AACACGAATG CGCCCCCAA GACCTGCAAG  
 551 CCCAATACGA CTGGGTAATC GACTGCCGGG GCTACGGCGC GAAAACCGCG  
 601 TGGAAACCAAT CCCCCGAGCA CACCAGCACC TTGCGCGGCA TACGCGGCGA  
 651 AGTGGCGCGG GTTTACACGC CCGAAATCAC GCTCAACCGC CCCGTGCGCC  
 701 TGCTGCACCC GCGCTATCCG CTCTACATCG CCCCAGAAAG AAACCACGTC  
 25 751 TTCGTCATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CCCCAGCAG  
 801 CGTACGTTCC GGGCTGGAAC TCTTATCCGC GCTCTATGCC GTCCACCCCG  
 851 CCTTCGGCGA AGCCGACATC CTCGAAATCG CCGCCGGCCT GCGCCCCACG  
 901 CTCAACCACC ACAACCCCGA AATCCGCTAC AGCCGCGAAC GCCGCCTCAT  
 951 CGAAATCAAC GGCCTTTTCC GGCACGGCTT TATGATTTC CCGCCGTAA  
 1001 CCGCCGCGCG CGTCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGACGCG  
 30 1051 CCCGAACGTG ATGAAGAAAG CGGTTTGGCG TATATCGGAA GACAAGATTA  
 1101 A

This corresponds to the amino acid sequence <SEQ ID 818; ORF126ng-1>:

1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTROGEHA AAYVAAAMLA  
 51 PAEEAVEATP EVIRLGRQSI PLWRGIRCRL NTLTMMQENG SLIVWHGQDK  
 35 101 PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ  
 151 LDGRQILSAL ADALDELNVP CHWEHECAPQ DLQAQYDWVI DCRGYGAKTA  
 201 WNQSPEHTST LRGIRGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENHV  
 251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIAAGLRPT  
 301 LNHHNPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA  
 40 351 PERDEESGLA YIGRQD\*

ORF126ng-1 and ORF126-1 show 95.1% identity in 366 aa overlap:

	10	20	30	40	50	60
orf126-1.pep	MTRIAVLGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP					
45 orf126ng-1	:					
	10	20	30	40	50	60
	70	80	90	100	110	120
orf126-1.pep	EVVRLGRQSIPLWRGIRCRLNTHMTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI					
50 orf126ng-1						
	70	80	90	100	110	120
	130	140	150	160	170	180
orf126-1.pep	VRWRADDAIEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECVPE					
55 orf126ng-1	:					
	130	140	150	160	170	180
	190	200	210	220	230	240
orf126-1.pep	GLQAQYDWLIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLHPRYP					
60 orf126ng-1	:					
	190	200	210	220	230	240
	DLQAQYDWVIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLHPRYP					
65						

-450-

		250	260	270	280	290	300
	orf126-1.pep	LYIAPKENHVFVIGATQIESESQAPASVRSGLLELLSALYAIHPAFGEADILEIATGLRPT					
5	orf126ng-1	LYIAPKENHVFVIGATQIESESQAPASVRSGLLELLSALYAVHPAFGEADILEIAAGLRPT					
		250	260	270	280	290	300
	orf126-1.pep	LNHHNPEIRYNRRRLIEINGLFRHGMISPAVTAARLAVAFDGDAPERDKESGLA					
10	orf126ng-1	LNHHNPEIRYSRERRLIEINGLFRHGMISPAVTAARLAVAFDGDAPERDEESGLA					
		310	320	330	340	350	360
15	orf126-1.pep	YIRRQDX					
	orf126ng-1	YIGRQDX					

Furthermore, ORF126ng-1 shows homology to a putative *Rhizobium* oxidase flavoprotein:

	gi 2627327 (AF004408) putative amino acid oxidase flavoprotein [Rhizobium etli]
	Length = 327
20	Score = 169 bits (423), Expect = 3e-41
	Identities = 112/329 (34%), Positives = 163/329 (49%), Gaps = 25/329 (7%)
	Query: 3 RIAVLGGGLSGRLTALQLAEQGYQIELFDKGTROGEHXXXXXXXXXXXXXXXXXXXXX 62
	RI V G G++G A QL G+++ L ++ G
25	Sbjct: 2 RILVNGAGVAGLTVAWQLYRHGFRVTLAERAGTVGA-GASGFAGGMLAPWCERESAEPEV 60
	Query: 63 IRLGRQSIPLWRGIRCLNTLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEIVR 122
	+ LGR + W + G+L+V G+D F R G DE+
30	Sbjct: 61 LTLGRLAADWWEAA-----LPGHVHRRGTLLVAGGRDTGELDRFSRRTS-GWEWLDEVA- 113
	Query: 123 WRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQDL 182
	IA EP L GRF ++ E LD RQ L+ALA L++ + +
	Sbjct: 114 -----IAALEPDLAGRFRRALFFRQEAHLDPQALAAALAGLEDARMRLTLG---VVGES 165
35	Query: 183 QAQYDWVIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLLLHPRPLY 242
	+D V+DC G LRG+RGE+ V T E++L+RPVRLLLHPR+P+Y
	Sbjct: 166 DVDHDRVVDCTGAA-----QIGRLPGLRGVRGEMLCVETTESLSRPVRLLLHPRHPIY 218
40	Query: 243 IAPKENHVFVIGATQIESESQAPASVRSGLLELLSALYAVHPAFGEADILEIAAGLRPTLN 302
	I P++ + F++GAT IES+ P + RS +ELL+A YA+HPAFGEA + E AG+RP
	Sbjct: 219 IVPDKNRFMVGATMIESDDGGPITARSLMELNAAAYAMHPAFGEARVTETGAGVRPAYP 278
	Query: 303 HHNPEIRYSRERRLIEINGLFRHGMISP 331
	+ P R ++E R + +NGL+RHGF+++P
45	Sbjct: 279 DNLP--RVTQEGRTLHVNGLYRHGFLAP 305

This analysis suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 97

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

50 819>:

	1	ATGACTGATA	ATCGGGGGTT	TACGCTGGTT	GAATTAATAT	CAGTGGTCTT
	51	GATATTGTCT	GTAATTGCTT	TAATTGTTTA	TCCGAGCTAT	CGCAATTATG
	101	TTGAGAAAGC	AAAGATAAAT	GCAGTGC GGG	CAGCCTTGTT	AGAAAATGCA
	151	CATTTTATGG	AAAAGTTTAA	TCTGCAGAAT	GGGAGGTTTA	AACAAACATC
55	201	TACCAAGTGG	CCAAGTTTGC	CGATTAAAGA	GGCAGAAGGC	TTTGTATCC
	251	GTTTGAATGG	AATCGtCGCG	CGGG. .GCTT	TAGACAGTAA	ATTCATGTTG
	301	AAGCGCGTAG	CCATAGATAA	AGATAAAAAA	CCTTTTATTA	TTAAGATGAA
	351	TGAAAATCTA	GTAACCTTTA	aTTTGCAAGA	AGTCCGCCAG	TTCGTGTAGT
	401	GACGGGCTGG	ATTATTTTAA	AGGAAATGAT	AAGGACTGCA	AGTTACTTAA
60	451	GTAG				

-451-

This corresponds to the amino acid sequence <SEQ ID 820; ORF127>:

```

1   MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLEN
51  HFMEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIVA RXALDSKFML
101 KAVAIDKDKN PFIKMNENL VTFICKKSAS SCSDDLDFYK GNDKCKLLK
151 *

```

Further work revealed the following DNA sequence <SEQ ID 821>:

```

1   ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
51  GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAAAT GCAGTGC GGG CAGCCTTGTT AGAAAATGCA
151 CATTTTATGG AAAAGTTTGA TCTGCAGAAT GGGAGGTTTA AACAAACATC
201 TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
251 GTTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This corresponds to the amino acid sequence <SEQ ID 822; ORF127-1>:

```

1   MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLEN
51  HFMEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
101 AVAIDKDKNP FIIKMNENLV TFICKKSASS CSDGLDFYKG NDKDCKLLK*

```

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF127 shows 98.0% identity over a 150aa overlap with an ORF (ORF127a) from strain A of *N.*

*meningitidis*:

```

25      orf127.pep      10      20      30      40      50      60
      orf127a          MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLEN AHMEKFYLQN
      orf127a          MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLEN AHMEKFYLQN
      orf127.pep      70      80      90      100     110     120
      orf127a          GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR XALDSKFMLK KAVAIDKDKNP FIIKMNENL
      orf127a          GRFKQTSTKW PSLPIKEAEG FCIRLNGI-ARGALDSKFMLK KAVAIDKDKNP FIIKMNENL
35      orf127.pep      130     140     150
      orf127a          VTFICKKSASS CSDGLDFYKG NDKDCKLLKX
      orf127a          VTFICKKSASS CSDGLDFYKG NDKDCKLLKX
40      orf127.pep      120     130     140     150
      orf127a          VTFICKKSASS CSDGLDFYKG NDKDCKLLKX

```

The complete length ORF127a nucleotide sequence <SEQ ID 823> is:

```

1   ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
51  GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAAAT ACAGTGC GGG CAGCCTTGTT AGAAAATGCA
151 CATTTTATGG AAAAGTTTGA TCTGCAGAAT GGGAGATTGA AACAAACATC
201 TACCAATGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
251 GTTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
50  401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 824>:

```

1   MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLEN
51  HFMEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
101 AVAIDKDKNP FIIKMNENLV TFICKKSASS CSDGLDFYKG NDKDCKLLK*

```

-452-

ORF127a and ORF127-1 show 99.3% identity in 149 aa overlap:

```

      10      20      30      40      50      60
orfl27a.pep MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINTVRAALLEN AHFMEKFY LQN
|||||
5 orfl27-1 MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLEN AHFMEKFY LQN
      10      20      30      40      50      60

      70      80      90      100     110     120
orfl27a.pep GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFI IKMNENLV
|||||
10 orfl27-1 GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFI IKMNENLV
      70      80      90      100     110     120

      130     140     150
orfl27a.pep TFICKKSASSCS DGLDYFKGNDKDKCKLLKX
|||||
15 orfl27-1 TFICKKSASSCS DGLDYFKGNDKDKCKLLKX
      130     140     150

```

## 20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF127 shows 97.3% identity over a 150 aa overlap with a predicted ORF (ORF127ng) from *N.gonorrhoeae*:

```

orfl27.pep MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLEN AHFMEKFY LQN 60
|||||
25 orfl27ng MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAAFLENAHFMEKFY LQN 60

orfl27.pep GRFKQTSTKWPSLPIKEAEGFCIRLNGIARXALDSKFMLKAVAIDKDKNPFI IKMNENL 120
|||||
30 orfl27ng GRFKQTSTKWPSLPIKEAEGFCIRLNGI-ARGALDSKFMLKAVAIDKDKNPFI IKMNENL 119

orfl27.pep VTFICKKSASSCS DGLDYFKGNDKDKCKLLK 150
|||||
orfl27ng VTFICKKSASSCS DRLDYFKGNDKDKCKLLK 149

```

The complete length ORF127ng nucleotide sequence <SEQ ID 825> is:

```

35      1 ATGACTGATA ATCGGGGGTT TACTACTGGTT GAATTAATAT CAGTGGTCTT
      51 GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
      101 TTGAGAAAGC AAAGATAAAT GCACTGCGGG CAGCCTTGTT AGAAAATGCA
      151 CATTTTATGG AAAAGTTTAA TCTGCAGAAT GGGAGATTTA AACAAACATC
      201 TACCAAATGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
40      251 GTTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
      301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
      351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
      401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 826>:

```

45      1 MTDNRGFTLV ELISVVLILSVLALIVYPSY RNYVEKAKIN AVRAAFLENA
      51 HFMEKFY LQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
      101 AVAIDKDKNP FIIKMENLV TFICKKSASS CSDRLDYFKG NDKDKCKLLK*

```

ORF127ng and ORF127-1 show 100.0% identity in 149 aa overlap:

```

      10      20      30      40      50      60
50 orfl27-1.pep MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLEN AHFMEKFY LQN
|||||
orfl27ng-1 MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLEN AHFMEKFY LQN
      10      20      30      40      50      60

      70      80      90      100     110     120
55 orfl27-1.pep GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFI IKMNENLV
|||||
orfl27ng-1 GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFI IKMNENLV
      70      80      90      100     110     120
60

```

-453-

		130	140	150
orf127-1.pep	TFICKKSASSCS	DGLDYFKGNDKDKCKLLKX		
orf127ng-1	TFICKKSASSCS	DGLDYFKGNDKDKCKLLKX		
		130	140	150

This analysis, including the fact that the predicted transmembrane domain is shared by the meningococcal and gonococcal proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 10 Example 98

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 827>

	1	..GTGTCGCTGG	CTTCGGTGAT	TGCCTCTCAA	ATCTTCCTTT	ACGAAGATTT
	51	CAACCAAATG	CGGAAAACCG	GTGGAGCTAT	CTGCGGTTTT	CTTGTCCAAT
15	101	ATTTATCTGG	GGTTTCAGCA	GGGGTATTTT	GATTTGAGTG	CCGACGAGAA
	151	CCCCGTACTG	CATATCTGGT	CTTTGGCAGT	AGAGGAACAG	TATTACCTCC
	201	TGTATCCCCT	TTTGCTGATA	TTTTGCTGCA	AAAAAACCAA	ATCGCTACGG
	251	GTGCTGCGTA	ACATCAGCAT	CATCCTGTTT	TTGATTTTGA	CTGCCTCATC
	301	GTTTTTGCCA	AGCGGGTTTT	ATACCGACAT	CCTCAACCAA	CCCAATACTT
	351	ATTACCTTTC	GACACTGAGG	TTTCCCGAGC	TGTTGGCAGG	TTCGCTGCTG
20	401	GCGGTTTACG	GGCAAACGCA	AAACGGCAGA	CGGCAAACAG	CAAATGGAAA
	451	ACGGCAGTTG	CTTTCATCAC	TCTGCTTCGG	CGCATTGCTT	GCCTGCCTGT
	501	TCGTGATTGA	CAAACACAAT	CCGTTTATCC	CGGGAATGAC	CCTGCTCCTT
	551	CCCTGCCTGC	TGACGGCACT	GCTTATCCGG	AGTATGCAAT	ACGGGACACT
	601	TCCGACCCGC	ATCCTGTCGG	CAAGCCCCAT	CGTATTTGTC	GGCAAATCT
25	651	CTTATCCCT	ATACCTGTAC	CATTGGATT	TTATTGCTTT	CGCTCCGCTC
	701	ATTAGAGGCG	GGAAACAGCT	CGGACTGCCT	GCCG..	

This corresponds to the amino acid sequence <SEQ ID 828; ORF128>:

	1	..VSLASVIASQ	IFLYEDFNQM	RKTVELSAVF	LSNIYLGFEQ	GYFDLSADEN
	51	PVLHIWSLAV	EEQYLLYPL	LLIFCCKTK	SLRVLRNISI	ILFLILTASS
30	101	FLPSGFYTDI	LNQPNYYLS	TLRFPELLAG	SLLAVYGQTQ	NGRRQTANGK
	151	RQLLSSLCFG	ALLACLFLVD	KHNPFIPIGT	LLLPCLLTAL	LIRSMQYCTL
	201	PTRILSASPI	VEVGKISYSL	YLYHWIFIAF	APLIRGGKQL	GLPA..

Further work revealed the complete nucleotide sequence <SEQ ID 829>:

	1	ATGCAAGCTG	TCCGATACAG	ACCGGAAATT	GACGGATTGC	GGGCCGTCGC
35	51	CGTGCTATCC	GTATGATTTT	TCCACCTGAA	TAACCGCTGG	CTGCCCGGAG
	101	GATTCCTGGG	GGTGACATT	TTCTTTGTCA	TCTCAGGATT	CCTCATTACC
	151	GGCATCATTC	TTTCTGAAAT	ACAGAACGGT	TCTTTTCTTT	TCCGGGATTT
	201	TTATACCCCG	AGGATTAAGC	GGATTATATC	TGCCTTTATT	GCGGGCGTGT
	251	CGCTGGCTTC	GGTGATTGCC	TCTCAAATCT	TCCTTTACGA	AGATTTC AAC
40	301	CAAATGCGGA	AAACCGTGGA	GCTTCTGCG	GTTTCTTGT	CCAATATTTA
	351	TCTGGGGTTT	CAGCAGGGGT	ATTTCGATT	GAGTGCCGAC	GAGAACCCCG
	401	TACTGCATAT	CTGGTCTTTG	GCAGTAGAGG	AACAGTATTA	CCTCCTGTAT
	451	CCCCTTTTGC	TGATATTTTG	CTGCAAAAAA	ACCAAATCGC	TACGGGTGCT
	501	GCGTAACATC	AGCATCATCC	TGTTTTTGAT	TTTGA CTGCC	TCATCGTTTT
45	551	TGCCAAGCGG	GTTTTATACC	GACATCCTCA	ACCAACCCAA	TACTTATTAC
	601	CTTTCGACAC	TGAGGTTTCC	CGAGCTGTTG	GCAGGTTGCG	TGCTGGCCGGT
	651	TTACGGGCAA	ACGCAAAACG	GCAGACGGCA	AACAGCAAAT	GGAAAACGGC
	701	AGTTGCTTTC	ATCACTCTGC	TTGCGCGCAT	TGCTTGCCCTG	CCTGTTCCGTG
	751	ATTGACAAAC	ACAATCCGTT	TATCCCGGGA	ATGACCCTGC	TCCTTCCCTG
50	801	CCTGCTGACG	GCACTGCTTA	TCCGAGTAT	GCAATACGGG	AACTTCCGA
	851	CCCGCATCCT	GTCGGCAAGC	CCCATCGTAT	TTGTCGGCAA	AATCTCTTAT
	901	TCCCTATACC	TGTACCATTG	GATTTTATT	GCTTTCGCCC	ATTACATTAC
	951	AGGCGACAAA	CAGCTCGGAC	TGCCTGCCGT	ATCGGCGGTT	GCCGCGTTGA
	1001	CGGCCGGATT	TTCCCTGTTG	AGTTATTATT	TGATTGAACA	GCCGCTTAGA
55	1051	AAACGGAAGA	TGACCTTCAA	AAAGGCATTT	TTCTGCCTCT	ATCTCGCCCC
	1101	GTCCCTGATA	CTTGTCGGTT	ACAACCTGTA	CGCAAGGGGG	ATATTGAAAC
	1151	AGGAACACCT	CCGCCCGTTG	CCCGGCGCGC	CCCTTGCTGC	GGAAAATCAT

5  
10  
15

```

1201 TTTCCGGAAG CCGTCCTGAC CCTCGGCGAC TCGCACGCCG GACACCTGAG
1251 GGGGTTTCTG GATTATGTCG GCAGCCGGGA AGGGTGGAAG GCCAAAATCC
1301 TGTCCCTCGA TTCGGAGTGT TTGGTTTGGG TAGATGAGAA GCTGGCAGAC
1351 AACCCGTTAT GTCGAAAATA CCGGGATGAA GTTGAAGAAAG CCGAAGCCGT
1401 TTTCATTGCC CAATTCATG ATTTGAGGAT GGGCGGCCAG CCTGTGCCGA
1451 GATTTGAAGC GCAATCCTTC CTAATACCCG GGTTCACGAG CCGATTGAGG
1501 GAAACCGTCA AAAGGATAGC CGCCGTCAA CCCGTCTATG TTTTGTGAAA
1551 CAACACATCA ATCAGCCGTT CGCCCTGAG GGAGGAAAAA TTGAAAAGAT
1601 TTGCCGCAAA CCAATATCTC CGCCCCATTC AGGCTATGGG CGACATCGGC
1651 AAGAGCAATC AGGCGGTCTT TGATTTGATT AAAGATATTC CCAATGTGCA
1701 TTGGGTGGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATATACG
1751 GCCGCTATCT TTACGGCGAC CAAGACCACC TGACCTATTT CGGTTCTTAT
1801 TATATGGGGC GGAATTTCCA CAAACACGAA CGCCTGCTTA AATCTTCCCA
1851 CGGCGGCGCA TTGCAGTAG

```

15 This corresponds to the amino acid sequence <SEQ ID 830; ORF128-1>:

20  
25

```

1 MQAVRYRPEI DGLRAVAVLS VMIFHLNNRW LPGGFLGVDI FFVISGFLIT
51 GIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN
101 QMRKTVELSA VFLSNIYLG FQQGYFDLSAD ENPVLHIWSL AVEEQYLLY
151 PLLLIFCCKK TKSLRVLRNI SIILFLILTA SSFLPSGFYT DILNQNTYY
201 LSTLRFPELL AGSLLAVYGQ TQNGRRQTAN GKRQLLSSLC FGALLACLFV
251 IDKHNPFIPG MTL LLPCLL ALLIRSMQYG TLPTRILSAS PIVFVGKISY
301 SLYLYHWIFI AFAHYITGDK QLGLPAVSAV AALTAFSLL SYLLIEQPLR
351 KRKMTFKKAF FCLYLAPSLI LVGYNLYARG ILKQEHRLPL PGAPLAAENH
401 FPETVLTLDG SHAGHLRGFL DYVGSREGWK AKILSLDSEC LVWVDEKLAD
451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPFRFAQSF LIPGFARFR
501 ETVKRIAIAVK PVYVFANNTS ISRSPLREEK LKRFAANQYL RPIQAMGDIG
551 KSNQAVFDLI KDIPNVHWVD AQKYLKNTV EIYGRYLYGD QDHLTYFGSY
601 YMGREFHKHE RLLKSSHGGA LQ*

```

Computer analysis of this amino acid sequence gave the following results:

30 Homology with hypothetical integral membrane protein HI0392 of *H.influenzae* (accession number U32723)

ORF128 and HI0392 show 52% aa identity in 180aa overlap:

35  
40

```

Orf128: 1 VSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG FQQGYFDLSADENPVLHIWSLAV 60
++L S IAS IF+Y DFN++RKT+EL+ FLSN YLG QGYFDLSA+ENPVLHIWSLAV
HI0392: 46 MALVSFIASAIIFYNDFNKLRTIELAIAFLSNFYLG LTQGYFDLSANENPVLHIWSLAV 105

Orf128: 61 EEQXXXXXXXXXIFCCKKTKSLRVLRNISIILFLILTASSFLPSGFYTDILNQNTYYLS 120
E Q I KK + ++VL I++ILF IL A+SF+ + FY ++L+QPN YYLS
HI0392: 106 EGQYYLIFPLILILAYKKFREVKVLFITLILFFILLATSFVSANFYKEVLHQPNYYLS 165

Orf128: 121 LTRFPELLAGSLLAVYGQTQNGRRQTANGKRQLLSSLCFGALLACLFVIDKHNPFIPGMT 180
LRFPELL GSLLA+Y N + Q + +L+ L L +CLF+++ + FIPG+T
HI0392: 166 NLRFPELLVGSLLAIYHNLSN-KVQLSKQVNNILAILSTLLLFSCFLMNNNIAFIPGIT 224

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

45 ORF128 shows 98.0% identity over a 244aa overlap with an ORF (ORF128a) from strain A of *N. meningitidis*:

50  
55  
60

```

orfl28.pep
10 20 30
VSLASVIASQIFLYEDFNQMRKTVELSAVF
|||||

orfl28a
60 70 80 90 100 110
ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVF

40 50 60 70 80 90
orfl28.pep
LSNIYLG FQQGYFDLSADENPVLHIWSLAVEEQYLLY PLLLIFCCKKTKSLRVLRNISI
|||||

orfl28a
120 130 140 150 160 170
LSNIYLG FQQGYFDLSADENPVLHIWSLAVEEQYLLY PLLLIFCCKKTKSLRVLRNISI

100 110 120 130 140 150
orfl28.pep
ILFLILTASSFLPSGFYTDILNQNTYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGK
|||||:|||||

```

-455-

orf128a		ILFLILTATSF	LP	SGFYTDIL	NQ	PNTYYL	STL	RFPELLAG	SLL	AVYGQT	QNGRRQT	TANGK
		180		190		200		210		220		230
5	orf128.pep	RQLLSSLC	FGALLAC	LFVIDKH	NPFI	PGMTLL	LPCLLT	TALLIR	SMQYGT	LPTRIL	SASPI	
		160		170		180		190		200		210
		240		250		260		270		280		290
		orf128a	RQLLSSLC	FGALLAC	LFVIDKH	NPFI	PGMTLL	LPCLLT	TALLIR	SMQYGT	LPTRIL	SASPI
10	orf128.pep	VFV	GKISYS	SLYLYH	WIFIAF	APLIR	GGKQL	GLPA				
		220		230		240						
		300		310		320		330		340		350
15	orf128a	VFV	GKISYS	SLYLYH	WIFIAF	AHYITG	DKQLGL	PAVSA	VAALTA	GFSL	SYLIEQ	PLRKR
		360		370		380		390		400		410
		orf128a	KMTF	KAFFCL	YLAPSL	ILVGYN	LYARGI	LKQEH	LRPLP	GAPLA	EAENH	FPETVLT
		360		370		380		390		400		410

The complete length ORF128a nucleotide sequence <SEQ ID 831> is:

20	1	ATGCAAGCTG	TCCGATACAG	ACCGGAAATT	GACGGATTGC	GGGCCGTCGC
	51	CGTGCTATCC	GTCATGATTT	TCCACCTGAA	TAACCGCTGG	CTGCCCGGAG
	101	GATTCTCTGGG	GGTGGACATT	TTCTTTGTCA	TCTCAGGATT	CCTCATTACC
	151	GGCATCATTC	TTTCTGAAAT	ACAGAACGGT	TCTTTTTTCTT	TCCGGGATTT
	201	TTATACCCGC	AGGATTAAGC	GGATTTATCC	TGCTTTTATT	GCGGCCGTGT
25	251	CGCTGGCTTC	GGTGATTGCC	TCTCAAATCT	TCCTTTACGA	AGATTTCAAC
	301	CAAATGCGGA	AAACCGTGGA	GCTTTCTGCG	GTTTTCTTGT	CCAATATTTA
	351	TCTGGGGTTT	CAGCAGGGGT	ATTTCGATTT	GAGTGCCGAC	GAGAACCCCG
	401	TACTGCATAT	CTGCTCTTTG	GCAGTAGAGG	AACAGTATTA	CCTCCTGTAT
	451	CCCTCTTTGC	TGATATTTTG	CTGCAAAAAA	ACAAAATCGC	TACGGGTGCT
	501	GCGTAACATC	AGCATCATCC	TATTTCTGAT	TTTGACTGCC	ACATCGTTTT
30	551	TGCCAAGCGG	GTTTTATACC	GATATTCTCA	ACCAACCCAA	TACTTATTAC
	601	CTTTCGACAC	TGAGGTTTCC	CGAGCTGTTG	GCAGGTTTCG	TGCTGGCGGT
	651	TTACGGGCAA	ACGCAAAACG	GCAGACGGCA	AACAGCAAAT	GGAAAACGGC
	701	AGTTGCTTTC	ATCACTCTGC	TTCGGCGCAT	TGCTTGCCCTG	CCTGTTCTGT
35	751	ATTGACAAAC	ACAATCCGTT	TATCCCGGGA	ATGACCCTGC	TCCTTCCCTG
	801	CCTGCTGACG	GCACTGCTTA	TCCGGAGTAT	GCAATACGGG	ACACTTCCGA
	851	CCCGCATCCT	GTCGGCAAGC	CCCATCGTAT	TTGTCGGCAA	AATCTCTTAT
	901	TCCCTATACC	TGTACCATTG	GATTTTTTAT	GCTTTCGCCC	ATTACATTAC
	951	AGGCGACAAA	CAGCTCGGAC	TGCCTGCCGT	ATCGGCGGTT	GCCGCGTTGA
40	1001	CGGCCGGATT	TTCCCTGTTG	AGTTATTATT	TGATTGAACA	GCCGCTTAGA
	1051	AAACGGAAGA	TGACCTTCAA	AAAGGCATTT	TTCTGCCTCT	ATCTCGCCCC
	1101	GTCCCTGATA	CTTGTCGGTT	ACAACCTGTA	CGCAAGGGGG	ATATTGAAAC
	1151	AGGAACACCT	CCGCCCCTTG	CCCGGCGCGC	CCCTTGCTGC	GGAAAATCAT
	1201	TTTCCGGAAA	CCGTCCTGAC	CCTCGGCGAC	TCGCACGCCG	GACACCTGCG
45	1251	GGGCTTCTG	GATTATTCG	GCAGCCGGGA	AGGGTGGA	AGGCTTCC
	1301	TGTCCTCGA	TTCGGAGTGT	TTGGTTTGGG	TAGATGAGAA	GCTGGCAGAC
	1351	AACCGTTAT	GTCGAAATA	CCGGGATGAA	GTTGAAAAG	CCGAAGCCGT
	1401	TTTCATTGCC	CAATTCTATG	ATTGAGGAT	GGGCGGCCAG	CCCGTGCCGA
	1451	GATTTGAAGC	GCAATCCTTC	CTAATACCCG	GGTTCACAG	CCGATTACAG
50	1501	GAAACCGTCA	AAAGGATAGC	CGCCGTCAAA	CCCGTCTATG	TTTTTGCAAA
	1551	CAACACATCA	ATCAGCCGTT	CGCCCTGAG	GGAGGAAAAA	TTGAAAAGAT
	1601	TTGCCGCAAA	CCAATATCTC	CGCCCCATTC	AGGCTATGGG	CGACATCGGC
	1651	AAGAGCAATC	AGGCGGTCTT	TGATTTGATT	AAAGATATTC	CCAATGTGCA
	1701	TGGGTGGAC	GCACAAAAT	ACCTGCCCAA	AAACACGGTC	GAAATATACG
55	1751	GCCGCTATCT	TTACGGCGAC	CAAGACCACC	TGACCTATTT	CGGTTCTTAT
	1801	TATATGGGGC	GGGAATTTCA	CAACACGAA	CGCCTGCTTA	AATCTTCTCG
	1851	CGACGCGCA	TTGCAGTAG			

This encodes a protein having amino acid sequence <SEQ ID 832>:

60	1	MQAVRYRPEI	DGLRAVAVLS	VMIFHLNNRW	LPGGFLGVDI	FFVISGFLIT
	51	GIILSEIQNG	SFSFRDFYTR	RIKRIYPAFI	AAVSLASVIA	SQIFLYEDFN
	101	QMRKTVELSA	VFLSNIIYLG	QQGYFDLSAD	ENPVLHIWSL	AVEEQYILLY
	151	PLLLIFCKCK	TKSLRVLRLNI	SIILFLILTA	TSFLPSGFYT	DILNQPNTRY
	201	LDLRFPELL	AGSLLAVYGQ	TQNGRRQTAN	GKRLSSLC	FGALLACLFV
	251	IDKHNPFI	PGMTLLPCLLT	ALLIRSMQYG	TLPTIRLSAS	PIVFGKISY
65	301	SLYLYHWIFI	AFAHYITGDK	QLGLPAVSAV	AALTAGFSL	SYLIEQPLR
	351	KRKMTFKKAF	FCLYLAPSLI	LVGYNLYARG	ILKQEHRLPL	PGAPLAENH
	401	FPETVLTG	LD SHAGHLRGFL	DYVGSREGWK	AKILSLDSEC	LVWVDEKLAD
	451	NPLCRKYRDE	VEKAEAVFIA	QFYDLRMGGQ	PVPRFEAQSF	LIPGFPARFR



501 ETVKRIAARK PVYVFANNTS ISRSPLREEK LKRFAANQYL RPIQAMGDIG  
 551 KSNQAVFDLI KDIPNVHWVD AQKYLKNTV EIYGRYLYGD QDHLTYFGSY  
 601 YMGREFHKHE RLLKSSRDGA LQ\*

ORF128a and ORF128-1 show 99.5% identity in 622 aa overlap:

5	orf128a.pep	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
	orf128-1	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
10	orf128a.pep	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
	orf128-1	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
15	orf128a.pep	QQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNISIILFLILTA
	orf128-1	QQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNISIILFLILTA
20	orf128a.pep	TSFLPSGFYTDILNQPNNTYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGKRQLSSLC
	orf128-1	SSFLPSGFYTDILNQPNNTYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGKRQLSSLC
25	orf128a.pep	FGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASP I VFVKISY
	orf128-1	FGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASP I VFVKISY
30	orf128a.pep	FCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLT LGDSHAGHLRGFL
	orf128-1	FCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLT LGDSHAGHLRGFL
35	orf128a.pep	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
	orf128-1	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
40	orf128a.pep	PVPRFEAQSF LIPGFPARFRET VKRIAARKPVYVFANNTSISRSPLREEK LKRFAANQYL
	orf128-1	PVPRFEAQSF LIPGFPARFRET VKRIAARKPVYVFANNTSISRSPLREEK LKRFAANQYL
45	orf128a.pep	RPIQAMGDIGKSNQAVFDLIKDI PNHWVDAQKYLKNTVEIYGRYLYGDQDHLTYFGSY
	orf128-1	RPIQAMGDIGKSNQAVFDLIKDI PNHWVDAQKYLKNTVEIYGRYLYGDQDHLTYFGSY
50	orf128a.pep	YMGREFHKHERLLKSSRDGALQX
	orf128-1	YMGREFHKHERLLKSSSHGGALQX

#### Homology with a predicted ORF from *N.gonorrhoeae*

50 ORF128 shows 93.4% identity over 244 aa overlap with a predicted ORF (ORF128ng) from *N. gonorrhoeae*:

	orf128.pep	VSLASVIASQIFLYEDFNQMRKTVELSAVF	30
	orf128ng	ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVF	112
55	orf128.pep	LSNIYLGFLQQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNISI	90
	orf128ng	LSNIYLGFLRGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCYKTKTKSLRVLNISI	172
60	orf128.pep	ILFLILTASSFLPSGFYTDILNQPNNTYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGK	150
	orf128ng	ILFLILTASSFLPAGFYTDILNQPNNTYYLSTLRFPPELLVGSLLAVYGQTQNGRRQTENGK	232
65	orf128.pep	RQLSSLCFGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI	210
	orf128ng	RQLSSLCFGALLVCLFVIDKHDPFIPGITLLLPCLLTALLIRSMQYGTLPTRILSASPI	292

orf128.pep VFVGKISYSLYLHWIFIAFAPLIRGGKQLGLPA 244  
 |||||  
 orf128ng VFVGKISYSLYLHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKR 352

5 The complete length ORF128ng nucleotide sequence <SEQ ID 833> is:

```

1 ATGCAAGCTG TCCGATACAG GCCTGAAATT GACGGATTGC GGGCCGTCGC
51 CGTGTATCC GTCATTATTT TCCACCTGAA TAACCGCTGG CTGCCCGGAG
101 GATTCTGGG GGTGGACATT TTCTTTGTCA TCTCGGGATT CCTCATTACC
151 AACATCATTC TTTCTGAAAT ACAGAACGGT TCTTTTCTT TCCGGGATTT
10 201 TTATACCCGC AGGATTAAGC GGATTTATCC TGCTTTTATT GCGGCCGTGT
251 CCCTGGCTTC GGTGATTGCT TCTCAAATCT TCCTTTACGA AGATTTCAAC
301 CAAATGAGGA AAACCATAGA GCTTCTACG GTTTTTTGT CCAATATTTA
351 TTTGGGGTTC CGATTGGGGT ATTTGATTG GAGTGCCGAC GAGAACCCCG
401 TACTGCATAT CTGGTCTTTG GCGGTAGAGG AACAGTATTA CCTCCTGTAT
15 451 CCTCTTTTGC TGATATTCTG TTACAAAAA ACCAAATCAC TACGGGTGCT
501 GCGTAATATC AGCATCATCC TGTTTCTGAT TTTGACCGCA TCATCGTTTT
551 TGCCGCGCCG GTTTTATACC GACATCCTCA ACCAACCcaa TACTTATTAC
601 CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GTGGGTTTCG TGTGGCGGT
651 TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGAAAAT GGAAAACGGC
20 701 AGTTGCTTTC ATTACTCTGT TTCGGCGCat tgCTTGTCTG CCTGTTCTGT
751 ATCGACAAAC ACGATCCGTT TATCCCGGGA ATAACCCTGC TCCTTCCCTG
801 CCTGTGACG GCGTGTCTTA TCCGGAGTAT GCAATACGGG ACACTTCCGA
851 CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT
901 TCCCTATACC TGTACCATTG GATTTTATTT GCCTTCGCC ATTACATTAC
25 951 AGGCGACAAA CAGCTCGGAC TGCTGCGGT ATCGGCGGT GCCGCGTTGA
1001 CGGCCGGATT TTCCCTGTTG AGCTATTATT TGATTGAACA GCCGCTTAGA
1051 AAACGGAAGA TGACCTTCAA AAAGGCATTT TTCTGCCTTT ATCTCGCCCC
1101 TGCCCTGATG CTTGTGCGTT ACAACCTGTA TTCAAGAGGG ATATTGAAAC
1151 AGGAACACCT CCGCCGCTG CCGGCGACGC CCGTTGCTGC GGAAAATAAT
30 1201 TTTCCGAAA CCGTCTTGAC CCTCGGCGAC TCGCACGCCG GACACCTGCG
1251 GGGGTTTCTG GATTATGTCG GCGGCAGGGA AGGGTGGAAA GCTAAAATCC
1301 TGTCCCTCGA TTCGGAGTGT TTGGTTTGGG TGGATGAGAA GCTGGCAGAC
1351 AACCCTTGT GCCGAAAATA CCGGGATGAA GTTGAAAAG CCGAAGCTGT
1401 TTTCAATTGCC CAATTCTATG ATTTGAGGAT GGGCGGCCAG CCCGTGCCGA
35 1451 GATTTGAAGC GCAATCCTTC CTGATACCCG GGTTCAAAGC CCGATTGAGG
1501 GAAACCGTCA AGAGGATAGC CGCCGTCAA CCTGTATATG TTTTGTCAA
1551 CAATACATCA ATCAGCCGTT CTCCTTGAG GGAGGAAAAA TTGAAAAGAT
1601 TTGCTATAAA CCAATACCTC CGGCCTATT GGGCTATGGG CGACATCGGC
1651 AAGAGCAATC AGCGGTCTT TGATTGGTT AAAGATATTC CCAATGTGCA
40 1701 TTGGGTGGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATACACG
1751 GACGCTATCT TTACGCGGAC CAAGACCACC TGACCTATTT CGGTTCTTAT
1801 TATATGGGGC GGAATTTCA CAAACACGAA CGCCTGCTCA AGCATCCCCG
1851 AGGCGGCGCA TTGCAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 834>:

```

45 1 MQAVRYRPEI DGLRAVAVLS VIIFHLNNRW LPGGFLGVDI FFVISGFLIT
51 NIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN
101 QMRKTIELST VFLSNIYLG FRLGYFDLSAD ENPVLHIWSL AVEEQYLLY
151 PLLIFCYKK TKSLRVL RNI SIILFLILTA SSFLPAGFYT DILNQPNITYY
50 201 LSTLRFPELL VGSLLAVYGQ TQNGRRQ TEN GKRQLLSLLC FGALLVCLFV
251 IDKHDPFIPG ITLLLPCLLT ALLIRSMQYG TLPTRILSAS PIVFVGKISY
301 SLYLHWIFI AFAHYITGDK QLGLPAVSAV AALTAGFSLL SYYLIEQPLR
351 KPMTFKKAF FCLYLAPSLM LVGYNLYSRG ILKQHLRPL PGT PVAENN
401 FRPVTLTGD SHAGHLRGFL DYVGGREGWK AKILSLDSEC LVWVDEKLAD
451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPRFEAQSF LIPGFKARFR
55 501 ETVKRIA AVK PVYVFANNTS ISRSPLREEK LKRFAINQYL RPIRAMGDIG
551 KSNQAVFDLV KDIPNVHVD AQKYL PKNTV EIHGRLYGD QDHLTYFGSY
601 YMGREFHKHE RLLKHSRGA LQ*

```

ORF128ng and ORF128-1 show 95.7% identity in 622 aa overlap:

```

60 orf128-1.pep MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
|||
orf128ng MQAVRYRPEIDGLRAVAVLSVIIFHLNNRWLPGGFLGVDIFFVISGFLITNIILSEIQNG

orf128-1.pep SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
|||
65 orf128ng SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVFLSNIYLG

```

```

5  orf128-1.pep  QQGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCCKKTKSLRVLNRNISIILFLILTA
   orf128ng      RLGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCYKKTSLRVLNRNISIILFLILTA

   orf128-1.pep  SSFLPSGFYTDILNQPNNTYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGKRQLLSSLC
   orf128ng      SSFLPAGFYTDILNQPNNTYYLSTLRFPPELLVGSLLAVYGQTQNGRRQTENGKRQLLSSLC

10  orf128-1.pep  FGALLACLFLVIDKHDPFIPGMITLLPCLLTALLIRSMQYGTLPTRILSASPPIVFGKISY
   orf128ng      FGALLVCLFLVIDKHDPFIPGITLLPCLLTALLIRSMQYGTLPTRILSASPPIVFGKISY

   orf128-1.pep  SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYLIEQPLRKRKMTFKKAF
   orf128ng      SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYLIEQPLRKRKMTFKKAF

   orf128-1.pep  FCLYLAPSLILVGYNLYARGILKQEHRLRPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL
   orf128ng      FCLYLAPSLMLVGYNLYSRGILKQEHRLRPLPGTPVAAENNFETVLTGLGDSHAGHLRGFL

20  orf128-1.pep  DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
   orf128ng      DYVGGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ

25  orf128-1.pep  PVPREFEASFLIPGFPPARFRETVKRIAARKPVYVFANNTSISRSPREEKLKRFAANQYL
   orf128ng      PVPREFEASFLIPGFKARFRETVKRIAARKPVYVFANNTSISRSPREEKLKRFAINQYL

30  orf128-1.pep  RPIQAMGDIGKSNQAVFDLIKDI PNHVWVDAQKYL PKNTVEIYGRYLYGDQDHLTYFGSY
   orf128ng      RPIRAMGDIGKSNQAVFDLVKDI PNHVWVDAQKYL PKNTVEIHGRYLYGDQDHLTYFGSY

   orf128-1.pep  YMGREFHKHERLLKSSHGGALQX
   orf128ng      YMGREFHKHERLLKHSRGGALQX
                        610      620

```

In addition, ORF218ng shows homology to a hypothetical *H. influenzae* protein:

```

40  sp|P43993|Y392_HAEIN  HYPOTHETICAL PROTEIN HI0392 >gi|1074385|pir||B64007
   hypothetical protein HI0392 - Haemophilus influenzae (strain Rd KW20)
   >gi|1573364 (U32723) H. influenzae predicted coding region HI0392 [Haemophilus
   influenzae] Length = 245
   Score = 239 bits (604), Expect = 3e-62
   Identities = 124/225 (55%), Positives = 152/225 (67%), Gaps = 1/225 (0%)

45  Query: 38  VDIFFVISGFLITNIILSEIQNGSFSFRDFYTRRIKRIYPXXXXXXXXXXXXXXXXXXFLYE 97
   +DIFFVISGFLIT II++EIQ SFS + FYTRRIKRIYP F+Y
   Sbjct: 1  MDIFFVISGFLITGIIITEIQNSFSLSKQFYTRRIKRIYPAFITVMALVSFIASAIIFYN 60

50  Query: 98  DFNQMRKTIELSTVFLSNIIYLGFRGLGYFDLSADENPVLHIWSLAVEEQXXXXXXXXXXIFC 157
   DFN++RKTIEL+ FLSN YLG GYFDLSA+ENPVLHIWSLAVE Q I
   Sbjct: 61  DFNKLRTIELAIAFLSNFYLGTLTQGYFDLSANENPVLHIWSLAVEGQYYLIFPLILILA 120

55  Query: 158 YKKTSLRVLNRNISIILFLILTASSFLPAGFYTDILNQPNNTYYLSTLRFPPELLVGSLLAV 217
   YKK + ++VL I++ILF IL A+SF+ A FY ++L+QPN YYLS LRFPELLVGSLLA+
   Sbjct: 121 YKKFREVKVLFIIITLILFILLATSFVSANFYKEVLHQPNIIYYLSNLRFPPELLVGSLLAI 180

   Query: 218 YGQTQNGRRQTENGKRQLLSLLCFGALLVCLFLVIDKHDPFIPGIT 262
   Y N + Q +L++L L CLF+++ + FIPGIT
60  Sbjct: 181 YHNLSN-KVQLSKQVNNILAILSTLLLFSCFLMNNNIAFIPGIT 224

```

This analysis, including the identification of several putative transmembrane domains, suggests that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 99**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 835>:

```

1  ..ATTATTTACG AATACCGCTG GATGTTTCTT TACGGCGCAC TGACGACCTT
51  GGGGCTGACG GTCGTGGCAA C.GCGGGCGG TTCGGTATG GGTCTGTGT
5  101  TGGCGTTGGC GCGCCTGATT CACTTGGAAA AAGCCGGTGC GCCGATGCGC
151  GTGCTGGCGT GGGCGTTGCG TAAAGTTTCG CTGCTGTATG TTACGCTGTT
201  CCGGGGTACG CCGCTGTTT TGCAGATTGT GATTGCGGCG TATGTGTGGT
251  TTCCGTTTTT CGTC..

```

This corresponds to the amino acid sequence <SEQ ID 836; ORF129>:

```

10  1  ..IIYEYRWMFL YGALTTLGLT VVAXAGGSVL GLLLALARLI HLEKAGAPMR
51  VLAWALRKVS LLYVTLFRGT PLFVQIVIWA YVWFPPFV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 837>:

```

1  ATGGATTTTC GTTTGACAT TATTTACGAA TACCGCTGGA TGTTTCTTTA
51  CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCAACG GCGGGCGGTT
15  101  CCGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGGAAAAA
151  GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AAGTTTCGCT
201  GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA
251  TTTGGGCGTA TGTGTGGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT
20  301  TTGGTCAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
351  GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
401  AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG
451  GCGCGTTCTT TGGGGCTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT
501  GCCGCAGGCA TTGCGCCGCA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA
25  551  CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GCGCGAGTTG
601  GCGTATGTTC AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC
651  GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
701  GGATATTCCT GCGTTTGAA AAACGTTACA ATCCGCAACA CCGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 838; ORF129-1>:

```

30  1  MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51  AGAPMRVLAW ALRKVSLLYV TFRGTPLEFV QIVIWAYVWF PFFVHPSDGI
101  LVSGEAAIAL RRGYGPLIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA
151  ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS EFITLLKDSS LLSVIAVAEL
201  AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

```

Computer analysis of this amino acid sequence gave the following results:

### 35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF129 shows 98.9% identity over a 88aa overlap with an ORF (ORF129a) from strain A of *N.*

*meningitidis*:

```

40  orf129.pep      10      20      30      40      50
      IIYEYRWMFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf129a  MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
      10      20      30      40      50      60

45  orf129.pep      60      70      80
      ALRKVSLLYVTLFRGTPLEFVQIVIWAYVWFPPFV
      ||||||||||||||||||||||||||||||||||||||
orf129a  ALRKVSLLYVTLFRGTPLEFVQIVIWAYVWFPPFVHPSDGILVSGEAAIALRRYGPLIAG
      70      80      90      100     110     120

50  orf129a  SLALIANSQAYICEIFRAGIQSIDKGQMEAARSLGLTYPQAMRYVILPQALRRMLPPLAS
      130     140     150     160     170     180

```

The complete length ORF129a nucleotide sequence <SEQ ID 839> is:

```

1  ATGGATTTTC GTTTGACAT TATTTACGAA TACCGCTGGA TGTTTCTTTA
51  CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCGACG GCGGGCGGTT

```

101 CGGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGGAAAAA  
 151 GCCGGTGGCG CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCGCT  
 201 GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA  
 251 TTTGGGCGTA TGTGTGGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT  
 5 301 TTGGTTAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT  
 351 GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG  
 401 AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG  
 451 GCGCGTTCTT TGGGGCTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT  
 501 GCCGCAGGCA TTGCGCCGTA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA  
 10 551 CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GGCGGAGTTG  
 601 GCGTATGTTC AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC  
 651 GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT  
 701 GGATATTCTT GCGTTTGAA AAACGTTACA ATCCGCAACA CCGCTGA

This encodes a protein having amino acid sequence <SEQ ID 840>:

15 1 MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK  
 51 AGAPMRVLAW ALRKVSLLYV TFRGTPLFV QIVIWAYVWF PFFVHPSDGI  
 101 LVSGEAAIAL RRGYGPLIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA  
 151 ARSLGLTYPQ AMRYVILPQA LRRMLPLAS EFITLLKDSS LLSVIAVAEL  
 201 AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR\*

20 ORF129a and ORF129-1 show 100.0% identity in 248 aa overlap:

orf129a.pep MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW  
 orf129-1 MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW  
 25 orf129a.pep ALRKVSLLYVTFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPLIAG  
 orf129-1 ALRKVSLLYVTFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPLIAG  
 30 orf129a.pep SLALIANSQAYICEIFRAGIQSIDKGQMEAARSLGLTYPQAMRYVILPQALRRMLPLAS  
 orf129-1 SLALIANSQAYICEIFRAGIQSIDKGQMEAARSLGLTYPQAMRYVILPQALRRMLPLAS  
 orf129a.pep EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE  
 35 orf129-1 EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE  
 orf129a.pep KRYNPQHRX  
 orf129-1 KRYNPQHRX  
 40

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF129 shows 98.9% identity over a 88 aa overlap with a predicted ORF (ORF129ng) from

*N.gonorrhoeae*:

45 orf129.pep IIYEYRWMFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW 54  
 orf129ng MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW 60  
 orf129.pep ALRKVSLLYVTFRGTPLFVQIVIWAYVWF PFFV 88  
 50 orf129ng ALRKVSLLYVTFRGTPLFVQIVIWAYVWF PFFVILHTAFLGNAMQSRVRPDKGRWIAG 120

An ORF129ng nucleotide sequence <SEQ ID 841> was predicted to encode a protein having amino acid sequence <SEQ ID 842>:

55 1 MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK  
 51 AGAPMRVLAW ALRKVSLLYV TFRGTPLFV QIVIWAYVWF PFFVILHTAF  
 101 LGNAMQSR RVPDKGRWIAG SLELNCQPRG RKTRGEFPPG ESNLGTEPRN  
 151 PLSMGQRREF GCENWYPPQN FIKK\*

Further work revealed the following gonococcal sequence <SEQ ID 843>:

1 ATGGATTTTc gtTTTGACAT TATTTAcgaA TACCGCTGGA TGTTTCTTTA

-461-

51 CCGCGCACTG Acgaccttgg ggctgacggt cgtggcgacg gCGGGCGGTT  
 101 CCGtattggG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTGGAAAAA  
 151 GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCGCT  
 201 GCTGTACGTT ACCCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA  
 251 TTTGGGCGTA TGTGTGGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT  
 301 TTGGTCAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT  
 351 GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG  
 401 AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG  
 451 GCGTGTCTT TGGGACTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT  
 501 CCCGCAGGCA TTGCGCCGTA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA  
 551 CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GCGGGAGTTG  
 601 GCGTATGTTT AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC  
 651 GCTTTACACC GCCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT  
 701 GGATATTCCT GCGTTTGAA AACGTTACA ATCCGCAACA CCGCTGA

15 This corresponds to the amino acid sequence <SEQ ID 844; ORF129ng-1>:

1 MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK  
 51 AGAPMRVLAW ALRKVSLLYV TLFRTPLFV QIVIWAYVWF PFFVHPSDGI  
 101 LVSGEAAIAL RRGYGPLIAG SLALIANSGA YICEIFRAGI QSIDKGQMEA  
 151 ARSLGLTYPQ AMRYVILPQA LRRMLPLAS EFITLLKDSS LLSVIAVAEL  
 201 AYWQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR\*

ORF129ng-1 and ORF129-1 show 99.2% identity in 248 aa overlap:

orf129-1.pep MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW  
 orf129ng-1 MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW  
 orf129-1.pep ALRKVSLLYVTTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPLIAG  
 orf129ng-1 ALRKVSLLYVTTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPLIAG  
 orf129-1.pep SLALIANSGAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPLAS  
 orf129ng-1 SLALIANSGAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPLAS  
 orf129-1.pep EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE  
 orf129ng-1 EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE  
 orf129-1.pep KRYNPQHRX  
 orf129ng-1 KRYNPQHRX

In addition, ORF129ng-1 is homologous to an ABC transporter from *A.fulgidus*:

2650409(AE001090) glutamine ABC transporter, permease protein (glnP)  
 [Archaeoglobus fulgidus] Length = 224  
 Score = 132 bits (329), Expect = 2e-30  
 Identities = 86/178 (48%), Positives = 103/178 (57%), Gaps = 18/178 (10%)  
 Query: 65 VSLLYVTTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPLIAGSLAL 124  
 +S YV + RGTPL VQI+I +F P+ GI + E A G +AL  
 Sbjct: 58 ISTAYVEVIRGTPLLVQILI-----VYFGLPAIGINLQPEPA-----GIIAL 99  
 Query: 125 IANSGAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPLASEFIT 184  
 SGAYI EI RAGI+SI GQMEAA SLG+TY QAMRYVI PQA R +LP L +EFI  
 Sbjct: 100 SICSGAYIAEIVRAGIESIPIGQMEAAARSLGMTYLQAMRYVIFPQAFRNILPALGNEFIA 159  
 Query: 185 LLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLEKR 242  
 LLKDSSLLSVI++ EL V I P AL YL+MT L + +K+  
 Sbjct: 160 LLKDSSLLSVISIVELTRVGRQIVNTTFNAWTPFLGVALFYLLMTIPLSRLVAYSQKK 217

This analysis, including the identification of transmembrane domains in the two proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 100**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 845>:

```

1   ..CTGAAAGAAT GCCGTCTGAA AGACCCTGTT TTTATTCCAA ATATCGTTTA
51  TAAGAACATC GCCATTACTT TCCTGCTCTT GCACGCCGCC GCCGAACTTT
5   101  GGCTGCCCCG GCAAACCGCC GGTTTTACCG CGCTCGCCGT CGGCTTCATC
151  CTGCTCGCCA AGCTGCGTGA gCTTCACCAT CACGAACCTT TACGTAAACA
201  cTACGTCCGC ACTTATTACy TGCTCCAAC TTTTGCCGCC GCAGgcTagT
251  TTGTGGACAG GCGCGGCGwA ATTACAAAC CTGCCCGCyT CCGCGCCCTT
301  GCACCTGATT ACCCTCGGCG GCATGATGGG CGGCGTGATG ATGGTGTGGc
10  351  TGACCGCCGG ACTGTGGCAC AGCGGCTTTA CCAAACCTGA CTACCCCAAA
401  CTCTGCCGCA TTGCCGTCCC CATCCTTTTC GCCGCCGCCG TCTCGCGCGC
451  TTTCTTGGrTG AACGTGAACC CGrTATTTT CATTACCGTT CCTGCGATTc
501  TGACCGCCGC CGTATTCTGA CTGTATCTTT TCrCGTTTAT ACCGATATTT
551  CGGCGCAATG CGTTTACAGA CGATCCGGAr TAR

```

15 This corresponds to the amino acid sequence <SEQ ID 846; ORF130>:

```

1   ..LKECRLKDPV FIPNIVYKNI AITFLLLHAA AELWLPAQTA GFTALAVGFI
51  LLAKLRELHH HELLRKHYVR TYLLQLFAA AGSLWTGAAX LQNLPASAPL
101  HLITLGGMMG GVMVWLTAG LWSHGFTKLD YPKLCRIAVP ILFAAAVSRA
151  FLXNVNPFXX ITVPAILTAA VFVLYLFXFI PIFRANAFTD DPE*

```

20 Further work revealed the complete nucleotide sequence <SEQ ID 847>:

```

1   ATGCGGCCGT TTTTCGTGCG CGCGGCGGTG CTTGCCATAC TCGGTGCGCT
51  GGTGTTTTTC ATCAACCCCG GTGCCATCGT CCTGCACCGC CAAATTTTCT
101  TGGAACCTAT GCTGCCGGCG GCATACGGCG GTTTTTTGAC TGCGGCTTTG
151  TTGGACTGGA CGGGTTTTTC GGGTAACCTG AAACCTGTCT CGACTTTGAT
25  201  GGCGGCATTA TTGCTCGCCG CATCCGCTAT ACTGCCCTTT TCGCCGCAAA
251  CTGCTCGTGT TTTCTGCGCC GCCTATTGGC TGGTGTGCT GCTGTTCTGC
301  GCGCGGCTGA TTTGGCTAGA CCGAAACACC GACAACTTCG CCCTGCTAAT
351  GTTACTTGCC GCGTTCAC TGTTTCAGAC GGCATATGCC GTCAGCGGCG
401  ATTTGAACCT GTTGCGCGCG CAAGTGCATC TAAATATGGC GGCGGTGATG
30  451  TTCGTATCCG TGCGCGTCAG TATTCTTTTG GGCGCGGAAG CCCTGAAAGA
501  ATGCCGTCTG AAAGACCTCG TTTTATTCC AAATATCGTT TATAAAAACA
551  TCGCCATTAC TTTCTGTCTC TTGCACGCCG CCGCCGAAC TGGGTGCCCC
601  GCGCAAACCG CCGGTTTTAC CGCGCTCGCC GTCGGCTTCA TCCTGCTCGC
651  CAAGCTGCGT GAGCTTCACC ATCAGAACT CTTACGTAAA CACTACGTCC
35  701  GCACTTATTA CTGCTCCAA CTCTTGCCG CCGCAGGCTA TTTGTGGACA
751  GCGCGGCGCA AATTACAAA CCTGCCCGCC TCCGCGCCCC TGCACCTGAT
801  TACCCTCGGC GGCATGATGG GCGGCGTGAT GATGGTGTGG CTGACCGCCG
851  GACTGTGGCA CAGCGGCTTT ACCAACTCG ACTACCCCA ACTCTGCCGC
901  ATTGCCGTCC CCATCCTTTT CGCCGCCGCC GTCTCGCGCG CTTTCTTGAT
40  951  GAACGTGAAC CCGATATTTT TCATTACCGT TCCTGCGATT CTGACCGCCG
1001 CCGTATTCGT ACTGTATCTT TTCACGTTA TACCGATATT TCGGGCGAAT
1051 GCGTTTACAG ACGATCCGGA ATAA

```

This corresponds to the amino acid sequence <SEQ ID 848; ORF130-1>:

```

1   MRPFVGAAGV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL
45  51  LDWTGFGSNL KPVATLMAAL LLAASAILPF SPQTASFFVA AYWLVLVLLFC
101  ARLIWLDNRN DNFALLMLLA AFTVEQTAYA VSGDLNLLRA QVHLNMAAVM
151  FVSVRVSILL GAEALKECRL KDPVFIPNIV YKNIAITFL LHAALWLP
201  AQTAGFTALA VGFILLAKLR ELHHHELLRK HYVRTYYLLQ LFAAAGYLWT
251  GAAKLQNLPA SAPLHLITLG GMMGVMMVW LTAGLWHS GF TKLDYPKLCR
50  301  IAVPILFAAA VSRAFLMNVN PIFITVPAI LTAAVFVLYL FTFIPIFRAN
351  AFTDDPE*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF130 shows 94.3% identity over a 193aa overlap with an ORF (ORF130a) from strain A of *N.*

55 *meningitidis*:

-463-

```

                                10      20      30
orf130.pep                      LKECRLKDPVFIPNIVYKNIAITFLLHAA
                                |||:|||:|||:|||:|||:|||:|||:|||
5  orf130a                      LNLRAQVHLNMAAVMFVSVRSILLGAEALKECRLKDPVFIPNVVYKNIAITFLLHAA
                                140      150      160      170      180      190

                                40      50      60      70      80      90
orf130.pep                      AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGSLWTGAAX
10 orf130a                      AELWLPAQTAGFTSLAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGYLWTGAAX
                                200      210      220      230      240      250

                                100      110      120      130      140      150
15 orf130.pep                      LQNLPAAPLHLITLGGMMGVMVWLTAGLWHSGF TKLDYPKLCRIAVPILFAAAVSRA
orf130a                      LQNLPAAPLHLITLGGMMGVMVWLTAGLWHSGF TKLDYPKLCRIAVPILFAAAVSRA
                                260      270      280      290      300      310

                                160      170      180      190
20 orf130.pep                      FLXNVNPFITVPAILTAADFVLYLFXFIPIFRANAFTDDPEX
orf130a                      VLMNVNPIFFITVPAILTAADFVLYLLTFVPIFRANAFTDDPEX
                                320      330      340      350

```

The complete length ORF130a nucleotide sequence <SEQ ID 849> is:

```

25      1  ATGCGGCCGT  TTTTCGTCGG  CGCGGCCGTG  CTTGCCATAC  TCGGTGCGCT
      51  GGTGTTTTTC  ATCAACCCCG  GTGCCATCGT  CCTGCACCGC  CAAATTTTCT
     101  TGGAACCTAT  GCTGCCGGCG  GCATACGGCG  GTTTTTTGAC  TCGCGCTTTG
     151  TTGGA CTGGA  CGGGTTTTTC  GGGTAACCTG  AAACCTGTCG  CGACTTTGAT
     201  GCGCGCATTA  TTGCTCGCCG  CATCCGCTAT  ACTGCCCTTT  TCGCCGCAAA
30     251  CTGCCTCGTT  TTTTCGTCGC  GCCTATTGGC  TGGTGTGCT  GCTGTTCTGC
     301  GCGCGCTGA  TTTGGCTAGA  CCGAAACACC  GACAACTTCG  CCCTGCTAAT
     351  GTTACTTGCC  GCGTTCAC TG  TTTTCAGAC  GGCAATATGCC  GTCAGCGGCG
     401  ATTTGAACCT  GTTGC GCGCG  CAAGTGCATC  TAAATATGGC  GCGGGTGATG
     451  TTCGTATCCG  TGCGCGTCAG  TATCTTTTGG  GGCGCGGAAG  CCCTGAAAGA
35     501  ATGCCGCTCG  AAAGACCCAG  TATTCATCCC  CAATGTCGTC  TATAAAAACA
     551  TCGCCATTAC  CTTCCTGCTC  CTGCACGCGG  CCGCCGAAC T  TGGCTGCCT
     601  GCGCAAACCG  CCGGTTTAC  CTCGCTCGCC  GTCGGCTTTA  TCCTGCTTGC
     651  CAAGCTGCGT  GAGCTTCACC  ATCAGCAACT  CCTGCGCAA  CACTACGTCC
40     701  GCACTTATTA  CCGTCTCAA  CTCTTGCCG  CCGCAGGCTA  TTTGTGGACA
     751  GCGCGCGCGA  AATTACAAA  CCTGCCCGCC  TCCGCGCCCC  TGCACCTGAT
     801  TACCCTCGGT  GGCATGATGG  GCAGCGTGAT  GATGGTGTGG  CTGACTGCCG
     851  GACTGTGCA  CAGCGGCTTT  ACCAAGCTCG  ACTACCCGAA  ACTCTGCCGC
     901  ATCGCCGTCC  CCATCCTNTT  CGCCGCGGCC  GTTTCGCGCG  CTGTTTAAAT
45     951  GAACGTAAAC  CCGATATCT  TCATACCGT  CCCGCAATT  CTGACCGCG
    1001  CCGTGTTCGT  GCTTTACCTG  CTGACATTCG  TACCGATCTT  TCGGGCGAAC
    1051  GCGTTTACAG  ACGATCCGGA  ATAA

```

This encodes a protein having amino acid sequence <SEQ ID 850>:

```

50      1  MRPFFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL
      51  LDWTGFSGNL KPVATLMAAL LLAASAILPF SPQTASFFVA AYWLVL LLLFC
     101  ARLIWLDRNT DNFALLMLLA AFTVFQTAYA VSGDLNLLRA QVHLNMAAVM
     151  FVSVRSILL GAEALKECRL KDPVFIPNVV YKNIAITFLL LHAAALWLP
     201  AQTAGFTSLA VGFILLAKLR ELHHHELLRK HYVRTYLLQ LFAAAGYLWT
     251  GAAKLQNLPA SAPLHLITLG GMMGSVMVW LTAGLWHS GF TKLDYPKLCR
55     301  IAVPILFAAA VSRVLMNVN PIFFITVPAI LTAADFVLYL LTFVPIFRAN
     351  AFTDDPE*

```

ORF130a and ORF130-1 show 98.3% identity in 357 aa overlap:

```

60  orf130a.pep      MRPFFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPAAYGGFLTAALLDWTGFSGNL
      |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
  orf130-1          MRPFFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPAAYGGFLTAALLDWTGFSGNL

  orf130a.pep      KPVATLMAALLLAASAILPFSPQTASFFVAAYWLVL LLLFCARLIWLDRNTDNFALLMLLA
      |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
  orf130-1          KPVATLMAALLLAASAILPFSPQTASFFVAAYWLVL LLLFCARLIWLDRNTDNFALLMLLA

65  orf130a.pep      AFTVFQTAYAVSGDLNLLRAQVHLNMAAVMFVSVRSILLGAEALKECRLKDPVFIPNVV

```



-464-

```

      |||
orfl30-1  AFTVFQTAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSILLGAEALKECRLKDPVFIPNIV
      |||
5  orfl30a.pep  YKNIAITFLLLHAAELWLPAQTAGFTSLAVGFILLAKLRELHHHELLRKHYVRTYYLLQ
      |||
orfl30-1  YKNIAITFLLLHAAELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQ
      |||
10 orfl30a.pep  LFAAAGYLWTGAAKLQNLPAAPLHLITLGGMMGSMVMVWLTAGLWHSFGFTKLDYPKLCR
      |||
orfl30-1  LFAAAGYLWTGAAKLQNLPAAPLHLITLGGMMGGVMVWLTAGLWHSFGFTKLDYPKLCR
      |||
15 orfl30a.pep  IAVPILFAAAVSRAVLNMVNPIFFITVPAILTAADFVLYLLTFVPIFRANAFTDDPE
      |||
orfl30-1  IAVPILFAAAVSRAFLNMVNPIFFITVPAILTAADFVLYLFTFIPIFRANAFTDDPE

```

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF130 shows 91.7% identity over a 193 aa overlap with a predicted ORF (ORF130ng) from *N.gonorrhoeae*:

```

20 orfl30.pep                                LKECRLKDPVFIPNIVYKNIAITFLLLHAA 30
      |||
orfl30ng  LNLRAQVHLNMAAVMFVSVRVSLLGTETLKECRLKDPVFIPNVIYKNIAIT-LLLHAA 201

25 orfl30.pep  AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGSLWTGAAX 90
      |||
orfl30ng  AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGYLWTGAAX 261

30 orfl30.pep  LQNLPAAPLHLITLGGMMGGVMVWLTAGLWHSFGFTKLDYPKLCRIAVPILFAAAVSRA 150
      |||
orfl30ng  LQNLPAAPLHLITLGGMTGGVMVWLTAGLWHSFGFTKLDYPKLCRIAVSILFASAVSRA 321

35 orfl30.pep  FLXNVNPFITVPAILTAADFVLYLFXFIPIFRANAFTDDPE 193
      |||
orfl30ng  VLMNVNPIFFITVPEILTAADFVLYLFTVPIFRANAFTDDPE 364

```

An ORF130ng nucleotide sequence <SEQ ID 851> was predicted to encode a protein having amino acid sequence <SEQ ID 852>:

```

40 1 MNKFFTHPMR PFFVGAAVLA ILGALVFFHQ PRRYHPAPPN FLGTYAAGCI
51 RRFDDYRFVG PDGFFRQPET CRYFDGGVVA CCGCFIAVET ATCRIFRRRL
101 LAGVA AVLRL ADLARRQHRT LRSVDVTA AF TVFQTAYAVS GDLNLLRAQV
151 HLNMAAVMFV SVRVS VLLGT ETLKECRLKD PVFIPNVIYK NIAITLLLHA
201 AEELWLPAQT AGFTALAVGF ILLAKLRELH HHELLRKHYV RTYYLLQLFA
251 AAGYLWTGAA KLQNLPAAP LHLITLGGMT GGVMVWLTG GLWHSFGFTKL
301 DYPKLCRIAV SILFASAVSR AVLNMVNPIF FITVPEILTA AVFMYLYLLTF
351 VPIFRANAFT DDPE*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 853>:

```

45 1 ATGCGCCCGT TTTTCGTCGG TCGGCAGTA CTTGCCATAC TCGGTGCGTT
51 GGTGTTTTTT ATCAACCCCG GCGCTATCAT CCTGCACCGC CAAATTTTCT
101 TGGAACCTAT GCTGCCGGCT GCATACGGCG GTTTTTTGAC TACCGCTTTG
151 TTGGACCGGA CGGTTTTTTC AGGCAACCTG AAACCTGCCG CTACTTTGAT
201 GGCGTGTTTG TTGCTGTGTTG CCGCTGTTTT ATTGCCGTTT TTACCGCAAC
50 251 TTGCCGCATT TTTCTGTCGC GCCTATTGGC TGGTGTGCT GCTGTTCTGC
301 GCCTGGCTGA TTTGGCTCGA CCGCAACACC GACAACTTCG CTCTGTTGAT
351 GTTACTTGCC GCATTTACCG TTTTTCAGAC GGCCTATGCC GTCAGCGGCG
401 ATTTGAACCT ACTGCGCGCG CAAGTGCATT TGAATATGGC GCGGTCATG
451 TTTGATCCG TCCGCTGTCAG CGTCCTTTTG GGCACGGAAA CCCTGAAAGA
55 501 ATGCCGCTCG AAAGACCCCG TATTCATCCC CAACGTTATC TATAAAACA
551 TCGCCATCAC CCTGCTGCTG CACGCCGCCG CCGAACTTTG GCTGCCCGCG
601 CAAACCGCCG GTTTTACTGC GCTTGCCGTC GGCTTCATCC TGCTCGCCAA
651 CACTCGCGAA CTGCACCATC ACGAATCTT ACGCAAACAC TACGTCCGCA
701 CTTATTACCT GCTCCAGCTC TTGCGCGCCG CAGGTTATCT GTGGACAGGC
60 751 GCGGCGAAAC TGCAAAACCT GCCCGCCTCC GCGCCCTGCG ACCTGATTAC
801 CCTCGCGGCG ATGACGGGTG GCGTGATGAT GGTGTGGCTG ACTGCCGGAC
851 TGTGGCACAG CGGCTTTACC AACTCGACT ACCCGAACT CTGCCGCATC

```

-465-

901 GCCGCTCCA TCCTTTTCGC CTCCGCCGTT TCGCGCGCTG TTTTAATGAA  
 951 CGTGAATCCG ATATTCTTCA TCACCGTTCC CGAGATTCTG ACCGCCGCCG  
 1001 TGTCATGCT TTACCTGCTG ACCTTCGTAC CGATTTTTCG AGCGAACGCC  
 1051 TTTACAGACG ATCCGGAATA A

5 This corresponds to the amino acid sequence <SEQ ID 854; ORF130ng-1>:

1 MRPFFVGA AV LAILGALVFF INPGAII LHR QIFLELMLPA AYGGFLTTAL  
 51 LDRTGFSGNL KPAATLMAVL LLVAAVLLPF LPQLAAFFVA AYWLVL L LFC  
 101 AWLIWLD RNT DNFALLMLLA AFTVFOTAYA VSGDLNLLRA QVHLNMAAVM  
 151 FVSVRVSVLL GTETLKECRL KDPVFIPNVI YKNIAITLLL HAAAEWLPA  
 10 QTAGFTALAV GFILLAKLRE LHHHELLRKH YVRTYYLLQL FAAAGYLWTG  
 251 AAKLQNL PAS APLHLITLGG MTGGVMMVWL TAGLWHS GFT KLDYPKLCRI  
 301 AVSILFASAV SRAVLMNVNP IFFITVPEIL TAAVFMLYLL TFVPIFRANA  
 351 FTDDPE\*

ORF130ng-1 and ORF130-1 show 92.4% identity in 357 aa overlap:

15	orf130-1.pep	MRPFFVGA AV LAILGALVFF INPGAII LHR QIFLELMLPA AYGGFLTTAL LDWTGFSGNL
	orf130ng-1	MRPFFVGA AV LAILGALVFF INPGAII LHR QIFLELMLPA AYGGFLTTAL LDRTGFSGNL
20	orf130-1.pep	KPVATLMA ALLLAASAILPFSPQTASFFVAAYWLVL L LFCARLIWLD RNTDNFALLMLLA
	orf130ng-1	KPAATLMA VLL LVA AVLLPFLPQLAAFFVAAYWLVL L LFCAWLIWLD RNTDNFALLMLLA
25	orf130-1.pep	AFTVFOTAYA VSGDLNLLRA QVHLNMAAVM FVSVRV SILLGAEALKECRLKDPVFIPNIV
	orf130ng-1	AFTVFOTAYA VSGDLNLLRA QVHLNMAAVM FVSVRV SVLLGTETLKECRLKDPVFIPNVI
30	orf130-1.pep	YKNIAITFLLHAAAEWLPA QTAGFTALAVGFILLAKLRELH HELLRKH YVRTYYLLQ
	orf130ng-1	YKNIAIT-LLHAAAEWLPA QTAGFTALAVGFILLAKLRELH HELLRKH YVRTYYLLQ
35	orf130-1.pep	LFAAGYLWTGAAKLQNL PASAPLHLITLGGMMGVMMVWL TAGLWHS GFTKLDYPKLCR
	orf130ng-1	LFAAGYLWTGAAKLQNL PASAPLHLITLGGMTGGVMMVWL TAGLWHS GFTKLDYPKLCR
40	orf130-1.pep	IAPVILFAAAVSRAFLMNVNPIFFITVPA ILTAAVFVLYLFTFIPIFRANAFTDDPEX
	orf130ng-1	IAVSILFASAVSRAVLMNVNPIFFITVPEIL TAAVFMLYLLTFVPIFRANAFTDDPEX

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 40 Example 101

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 855>:

1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCA TT  
 51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA  
 45 101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATT TTTGGATATT  
 151 GCGGCGGAGA GTCCGCCGTC TTTAGGGGAC TACGAGATAC CGCTTTCAGA  
 201 CCGCAATAGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT  
 251 ACTTTTACAG GAAATAGGG AAGTTGAAG C.TGCGGGCT GGATTGGCGT  
 301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTGGA  
 351 CTGCTTGGA AAG..

50 This corresponds to the amino acid sequence <SEQ ID 856; ORF131>:

1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI  
 51 GGESPPSLGD YEIPLSDGNS SVRANEYESA QQSYFYRKIG KFEXCGLDWR  
 101 TRDGKPLIET FKQGGFDCLE K..

Further work revealed the complete nucleotide sequence <SEQ ID 857>:

55 1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCA TT  
 51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA

-466-

5  
101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATT TTTGGGATATT  
151 GGCGGCGAGA GTCCGCCGTC TTTAGGGGAC TACGAGATAC CGCTTTTCAGA  
201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT  
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGCT GGATTGGCGT  
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTGA  
351 CTGCTTGGA AAGCAGGGGT TCGGCGCAA CCGTCTGTCC GAGCGCGTCC  
401 GATGGTAA

This corresponds to the amino acid sequence <SEQ ID 858; ORF131-1>:

10  
1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI  
51 GGESPPSLGD YEIPLSDG NR SVRANEYESA QSYFYRKIG KFEACGLDWR  
101 TRDGKPLIET FKQGGFDCLE KQGLRRNGLS ERVRW\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF131 shows 95.0% identity over a 121aa overlap with an ORF (ORF131a) from strain A of *N.*

15 *meningitidis*:

		10	20	30	40	50	60
orfl31.pep		MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD					
orfl31a		MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLED					
		10	20	30	40	50	60
orfl31.pep		70	80	90	100	110	120
orfl31a		YEIPLSDGNSSVRANEYESAQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCLE					
		70	80	90	100	110	120
orfl31a		YEIPLSDGNRSVRANEYESAQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLE					
orfl31.pep	K						
orfl31a	KQGLRRNGLSERVRWX						
	130						

The complete length ORF131a nucleotide sequence <SEQ ID 859> is:

35  
1 ATGGAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCAAT  
51 TACGGTTGCA GGCTGCCGGT TGGCAGGTTG GTATAGAGTG TCGTCCCTGT  
101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATT TTTGGGATATT  
151 GGCGGCGAGA GTCCCTCCGTC TTTAGAGGAC TACGAGATAC CGCTTTTCAGA  
201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT  
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGTT GGATTGGCGT  
40  
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG AAGGTTTGA  
351 TTGTTTGA AAGCAGGGGT TCGGCGCAA CCGTCTGTCC GAGCGCGTCC  
401 GATGGTAA

This encodes a protein having amino acid sequence <SEQ ID 860>:

45  
1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI  
51 GGESPPSLED YEIPLSDG NR SVRANEYESA QSYFYRKIG KFEACGLDWR  
101 TRDGKPLIET FKQEGFDCLE KQGLRRNGLS ERVRW\*

ORF131a and ORF131-1 show 97.0% identity in 135 aa overlap:

orfl31a.pep	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLED
orfl31-1	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD
orfl31a.pep	YEIPLSDGNRSVRANEYESAQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLE
orfl31-1	YEIPLSDGNRSVRANEYESAQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCLE
orfl31a.pep	KQGLRRNGLSERVRWX

orf131-1 KQGLRRNGLSERVRWX

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF131 shows 89.3% identity over 121 aa overlap with a predicted ORF (ORF131ng) from

5 *N.gonorrhoeae*:

```

orf131.pep  MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD  60
            ||||:|||| |||:||||||| |||:||||||| |||:||||||| |||
orf131ng    MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPLSLED  60

10 orf131.pep  YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCLE  120
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf131ng    YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE  120

15 orf131.pep  K 121
            |
orf131ng      KQGLRRNGLSERVRW 134

```

A complete length ORF131ng nucleotide sequence <SEQ ID 861> was predicted to encode a protein having amino acid sequence <SEQ ID 862>:

```

20      1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC LSLSGWCKPR KPAAIDFWDI
      51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
     101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 863>:

```

25      1 ATGGAATTC GGGTAATAAA ATATACGGCA ACGGCTGCGT TGTTCGCATT
      51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCTTGT
     101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATTT TTGGGATATT
     151 GGCGGCGAGA GtccgctGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA
     201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCG CAAAAATCTT
     251 ACTTTTATAG GAAAATAGGG AAGTTTGAAG CCTGCGGGTT GGATTGCGGT
     301 ACGCGTGACG GCAAACCTTT GGTGAGAGG TTCAAACAGG AAGGTTTCGA
30  351 CTGTTTGGA AAGCAGGGGT TGCGGCGCAA CGGCCTGTCC GAGCGCGTCC
     401 GATGGTAA

```

This corresponds to the amino acid sequence <SEQ ID 864; ORF131ng-1>:

```

35      1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI
      51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
     101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*

```

ORF131ng-1 and ORF131-1 show 92.6% identity in 135 aa overlap:

```

40 orf131ng-1.pep MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPLSLED
            ||||:|||| |||:||||||| |||:||||||| |||:||||||| |||
orf131-1      MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD

orf131ng-1.pep YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf131-1      YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCLE

45 orf131ng-1.pep KQGLRRNGLSERVRWX
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf131-1      KQGLRRNGLSERVRWX

```

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 102**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 865>

```

      1  ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
    51  TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
  101  AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
  151  TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
  201  CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
  251  TGAACCTCGG CTGCCTTAT ATTtCGGCC CGCAATGGCT GTCGGAAGAAC
  301  GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACgC ACGGCAAAAC
  351  GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATgCC GGCTCGCGC
  401  CGGGCTTCCT TATtGGCGGC GTACC.GGAA AATtCGGCG TTTCCGCCCG
  451  CCTGCCGCAA ACGCCGCGCC AAGACCGGAA CAGCCAATCG CCGTTTTTcG
  501  TCATCGAAGC CGACGAATAC GACACCGCCT TTtTCGACAA ACGTTCTAAA
  551  TtCGTGCATT ACCGTCCGCG TACCGCCGTG TTGAACAATC TGGAAATCGA
  601  CCACGCCGAC ATCTTTGCCG ACTTGGGCGC GATACAGACc CAGTTCCACT
  651  ACCTCGTGGC TACCGTGCCG TCTGAAGGCT TAATCGTCTG CAACGGACGG
  701  CAGCAAAGCC TGCAAGATAC TTTGGACAAA GGCTGCTGGA CGCCGGTGGA
  751  AAAATTTCGC ACGGAACACG GCTGGCA..

```

This corresponds to the amino acid sequence <SEQ ID 866; ORF132>:

```

  20      1  MKHIHIIGIG GTFMGGGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
    51  YEGFDAQLD EFKADVYVIG NVAKRGM DVV EAILNLGLPY ISGPQWLSEN
  101  VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VXGKERRFRP
  151  PAANAAPRPE QPIAVFRHRS RRIRHRLFRQ TFXIRALPSA YRRVEQSGIR
  201  PRRHLCLRGR DTDVPLPRA YRAVXRLNRL QRTAAKPARY FGQRLLDAGG
  251  KIRHGTRLA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 867>:

```

      1  ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
    51  TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
  101  AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
  151  TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
  201  CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
  251  TGAACCTCGG CTGCCTTAT ATTTCCGGCC CGCAATGGCT GTCGGAAGAAC
  301  GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACGC ACGGCAAAAC
  351  GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATGCC GGCTCGCGC
  401  CGGGCTTCCT TATTGGCGGC GTACCGGAAA ATTTCCGGCT TCCGCCCGC
  451  CTGCCGCAA CGCCGCGCCA AGACCGGAAC AGCCAATCGC CGTTTTTCGT
  501  CATCGAAGCC GACGAATACG ACACCGCCTT TTTCGACAAA CGTTCTAAAT
  551  TCGTGCATTA CCGTCCGCGT ACCGCGGTGT TGAACAATCT GGAATTCGAC
  601  CACGCCGACA TCTTTGCCGA CTGGGCGCG ATACAGACCC AGTTCCACTA
  651  CCTCGTGCGT ACCGTGCCGT CTGAAGGCTT AATCGTCTGC AACGGACGGC
  701  AGCAAAGCCT GCAAGATACT TTGGACAAAG GCTGCTGGAC GCCGGTGGAA
  751  AAATTCGGCA CGGAACACGG CTGGCAGGCC GGCGAAGCCA ATGCCGACGG
  801  CTCGTTTCGAC GTGTTGCTCG ACGGCAAAAC CGCCGGACGC GTCAAATGGG
  851  ATTTGATGGG CAGGCACAAC CGCATGAACG CGCTCGCCGT CATTGCCGCC
  901  GCGGTCATG TCGGTGTCGA TATTCAGACC GCCTGCGAAG CCTTGGGCGC
  951  GTTTAAAAAC GTCAAACGCC GGATGGAAT CAAAGGCACG GCAAACGGCA
  1001 TCACCGTTTA CGACGACTTC GCCCACCACC CGACCGCCAT CGAAACCACG
  1051 ATTCAGGTT TGCGCCAACG CGTCGGCGGC GCGCGCATCC TCGCCGTCTT
  1101 CGAACCGCGT TCCAACACGA TGAAGCTGGG CACGATGAAG TCCGCCCTGC
  1151 CTGTAAGCCT CAAAGAAGCC GACCAAGTGT TCTGCTACGC CGGCGGCGTG
  1201 GACTGGGACG TCGCCGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGAACGT
  1251 CGGCAAAGAC TTCGATGCCT TCGTTGCCGA AATCGTGAAG AACGCCGAAG
  1301 TAGGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC
  1351 GGAAAGCTGC TGGAAGCTTT GAGATAG

```

55 This corresponds to the amino acid sequence <SEQ ID 868; ORF132-1>:

```

      1  MKHIHIIGIG GTFMGGGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
    51  YEGFDAQLD EFKADVYVIG NVAKRGM DVV EAILNLGLPY ISGPQWLSEN
  101  VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VPENFGVSAR
  151  LPQTPRQDPN SQSPFFVIEA DEYDTAFFDK RSKFVHYRPR TAVLNNLEFD
  201  HADIFADLGA IQTQFHYLVR TVPSEGLIVC NGRQQSLQDT LDKGCWTPVE
  251  KFGTEHWQA GEANADGSFD VLLDGKTAGR VKWDLMGRHN RMNALAVIAA
  301  ARHVGVDIQT ACEALGAFKN VKRRMEIKGT ANGITVYDDF AHHPTAIETT

```

351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK SALPVSLKEA DQVFCYAGGV  
401 DWDVAEALAP LGGRLNVGKD FDAFVAEIVK NAEVGDHILV MSGGGFGGGIH  
451 GKILLEALR\*

Computer analysis of this amino acid sequence gave the following results:

5 Homology with the hypothetical o457 protein of *E.coli* (accession number U14003)

ORF132 and o457 show 58% aa identity in 140 aa overlap:

	Orf132:	4	IHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLDEFK	63
			IHI+GI GTFMGGGLA +A++ G EV+G DA +YPPMST LE GI++ +G+DA+QL+ +	
10	o457:	3	IHILGICGTFMGGGLAMLARQLGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-Q	61
	Orf132:	64	ADVYVIGNVAKRGM DVVEAILNLGLPYISGPQWLSENVLHHHVVVLGVAGTHGKTTTASML	123
			D+ +IGN RG VEA+L +PY+SGPQWL + VL WVL VAGTHGKTTTA M	
	o457:	62	PDLVIIGNAMTRGNPCVEAVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMA	121
15	Orf132:	124	AWVLEYAGLAPGFLIGGVXG	143
			W+LE G PGF+IGGV G	
	o457:	122	TWILEQCGYKPGFVIGGVPG	141

### Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF132 shows 74.6% identity over a 189aa overlap with an ORF (ORF132a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
25	orf132.pep	MKHIHIIGIGGTFMGG	LAAIAKEAGFEVSG	DAKMYPPMSTQLEAL	GIDVYEGFDA	QQLD	
	orf132a	MKHIHIIGIGGTFMGG	IAAIAKEAGFEXSG	DAKMYPPMSTQLEAL	GIGVYEGFDTA	QLD	
		10	20	30	40	50	60
30	orf132.pep	70	80	90	100	110	120
		EFKADVYVIGNVAKR	GMDVVEAILNLGL	PYISGPQWLSENV	LHHHWLVGVAG	THGKTTTA	
	orf132a	EFKADVYVIGNVAKR	GMDVVEAILNRGL	PYISGPQWLAENX	LHHHWLVGVAX	THGKTTTA	
		70	80	90	100	110	120
35	orf132.pep	130	140	150	160		
		SMLAWVLEYAGLAPG	FLIGGVXGKFR---	RFRPPAANAAPR	PEQPI-----	AVFR	
				:	: :	:: :	
	orf132a	SMLAWVLEYAGLAPG	FXIGGVPE NFSV	SARL-PQTPRQDP	NSQSPFFVIEA	EYDTAFFD	
		130	140	150	160	170	
40	orf132.pep	170	180	190	200	210	220
		HRSRRIRHRLFRQT	FXIRALPSAYRR	VEQSGIRPRRH	LCRLGRDTPV	PLPRAYRAV	XRL
		:  :	:::				
	orf132a	KRSKFVHYRPRTA	VLNNLEFDHAD	IFADLGAIQTQ	FHHLVRTVPSE	GLIVCNGRQ	QSLQD
		180	190	200	210	220	230

45 The complete length ORF132a nucleotide sequence <SEQ ID 869> is:

	1	ATGAAACACA	TCCACATTAT	CGGTATCGGC	GGCACGTTTA	TGGGTGGGAT
	51	TGCCGCCATT	GCCAAAGAAG	CAGGGTTTGA	ANTCAGCGGT	TGCGATGCGA
	101	AGATGTATCC	GCCGATGAGC	ACCCAGCTCG	AAGCCTTGGG	CATAGGCGTG
50	151	TATGAAGGCT	TCGACACCGC	GCAGTTGGAC	GAATTTAAAG	CCGACGTTTA
	201	CGTTATCGGC	AATGTCCGCA	AGCGCGGGAT	GGATGTGGTT	GAAGCGATTT
	251	TGAACCGTGG	GCTGCCTTAT	ATTTCCGGCC	CGCAATGGCT	GGCTGAAAAC
	301	NTGCTGCACC	ATCATTGGNN	ACTCGGCGTG	GCGNGACGC	ACGGCAAAC
	351	GACCACCGCG	TCTATGCTCG	CGTGGGTTTT	GGAAATATGCC	GGACTCGCAC
55	401	CGGGCTTCNT	TATCGGCGGC	GTACCGGAA	ACTTCAGCCT	TTCCGCCCCG
	451	CTGCCGCAAA	CGCCCGCCCA	AGACCCGAAC	AGCCAATCGC	CGTTTTTCGT
	501	CATTGAAGCC	GACGAATACG	ACACCGCGTT	TTTCGACAAA	CGCTCCAAAT
	551	TCGTGCATTA	CCGTCCGCGT	ACCGCCGTGT	TGAACAATCT	GGAATTCGAC
	601	CACGCCGACA	TCTTCGCCGA	TTTGGGCGCG	ATACAGACCC	AGTTCACACA
60	651	CCTCGTGCGT	ACCGTGCCGT	CTGAAGGCCT	CATCGTCTGC	AACGGACCGC
	701	AGCAAAGCCT	GCAAGACACT	TTGGACAAAG	GCTCTGGAC	CCCGGTGGAA
	751	AAATTCCGCA	CGGAACACGG	CTGGCAGGCC	GGCGAAGCCA	ATGCCGATGG

-470-

801 CTCGTTTCGAC GTGTTGCTTG ACGGCAAAAA AGCCGGACAC GTCGCTTGA  
 851 GTTTGATGGG CGGACACAAC CGCATGAACG CGCTCGCNGT CATCGCCGCC  
 901 GCGCGTCATG CCGGAGTNGA CATTGAGACG GCCTGCGAAG CCTTGAGCAC  
 951 GTTTAAAAAC GTCAAACGCC GCATGGAAT CAAAGGCACG GCAAACGGTA  
 1001 TCACCGTTTA CGACGACTTC GCCACCATC CGACCGCTAT CGAAACCACG  
 1051 ATTCAAGGTT TGCGCCAGCG CGTCGGCGGC GCGCGCATCC TCGCCGTCCT  
 1101 CGAACCGCGT TCCAATACGA TGAAGCTGGG TACGATGAAA GCCGCCCTGC  
 1151 CCGCAAGCCT CAAAGAAGCC GACCAAGTGT TCTGNTACGC CGGCGGCGCG  
 1201 GACTGGGACG TTGCCGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGCACGT  
 1251 CGGCAAAGAC TCGATGCCT TCGTTGCCGA AATCGTGAAA AACGCCGAAG  
 1301 CAGGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC  
 1351 ACCAACTGC TGGACGCTTT GAGATAG

This encodes a protein having amino acid sequence <SEQ ID 870>:

1 MKHIHIIGIG GTFMGGIAAI AKEAGFEXSG CDAKMYPPMS TQLEALGIGV  
 51 YEGFDTAQLD EFKADVYVIG NVAKRGMDVV EAILNRGLPY ISGPQWLAEN  
 101 XLHHHWXLGV AXTHGKTTTA SMLAWVLEYA GLAPGFIXIG VPENFVSAR  
 151 LPQTPRQDPN SQSPFFVIEA DEYDTAFFDK RSKFVHYRPR TAVLNNLEFD  
 201 HADIFADLGA IQTQFHHLVR TVPSEGLIVC NGRQQSLQDT LDKGCWTPVE  
 251 KFGTEHGWQA GEANADGSFD VLLDGKKAGH VAWSLMGGHN RMNALAVIAA  
 301 ARHAGVDIQT ACEALSTFKN VKRRMEIKGT ANGITVYDDF AHHPTAIETT  
 351 IQGLRQRVGG ARILAVLEPR SNTMKLGMTK AALPASLKEA DQVFXYAGGA  
 401 DWDVAEALAP LGGRLHVGKD FDAFVAEIVK NAEAGDHILV MSNGGFGGIH  
 451 TKLLDALR\*

ORF132a and ORF132-1 show 93.9% identity in 458 aa overlap:

25 orf132a.pep MKHIHIIGIGGTFMGGIAAIAKEAGFEXSGCDAKMYPPMSTQLEALGIGVYEGFDTAQLD  
 orf132-1 MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD  
 30 orf132a.pep EFKADVYVIGNVAKRGMDVVEAILNRGLPYISGPQWLAENXLHHHWXLGVAXTHGKTTTA  
 orf132-1 EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA  
 orf132a.pep SMLAWVLEYAGLAPGFIXIGGVPEFVSARLPQTPRQDPNSQSPFFVIEADEYDTAFFDK  
 35 orf132-1 SMLAWVLEYAGLAPGFLIGGVPEFVSARLPQTPRQDPNSQSPFFVIEADEYDTAFFDK  
 orf132a.pep RSKFVHYRPR TAVLNNLEFDHADIFADLGA IQTQFHHLVRTVPSEGLIVC NGRQQSLQDT  
 40 orf132-1 RSKFVHYRPR TAVLNNLEFDHADIFADLGA IQTQFHYLVRTVPSEGLIVC NGRQQSLQDT  
 orf132a.pep LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKKAGHVAWSLMGGHNRMNALAVIAA  
 orf132-1 LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKTAGRVKWDLMGRHNRNALAVIAA  
 45 orf132a.pep ARHAGVDIQTACEALSTFKNVKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG  
 orf132-1 ARHVGVDIQTACEALGAFKNVRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG  
 50 orf132a.pep ARILAVLEPR SNTMKLGMTKAALPASLKEADQVFXYAGGADWDVAEALAPLGGRLHVGKD  
 orf132-1 ARILAVLEPR SNTMKLGMTKSALPVSLKEADQVFCYAGGVDWDVAEALAPLGGRLNVGKD  
 orf132a.pep FDAFVAEIVKNAEAGDHILVMSNGGFGGIHTKLLDALRX  
 55 orf132-1 FDAFVAEIVKNAEAGDHILVMSNGGFGGIHTKLLDALRX

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF132 shows 89.6% identity over 259 aa overlap with a predicted ORF (ORF132ng) from *N. gonorrhoeae*:

60 orf132.pep MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD 60  
 orf132ng MKHIHIIGIGGTFMGGIAAIAKEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAAQLE 60

-471-

	orf132.pep	EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA	120
	orf132ng	EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAENVLHHHWLVGVAGTHGKTTTA	120
5	orf132.pep	SMLAWVLEYAGLAPGFLIGGVXGKFRFRPPAANAAPRPEQPIAVFRHRSRRIRHRLFRQ	180
	orf132ng	SMLAWVLEYAGLAPGFLIGGVPGKFRFRPPAANAASRPEQPIAVFRHRSRRIRHRLFRQ	180
10	orf132.pep	TFXIRALPSAYRRVEQSGIRPRRHLCRLGRDTPVPLPRAYRAVXRLNRLQRTAAKPARY	240
	orf132ng	TLQIRALSPAYRRVEQSGIRPRRHRLRLGRDTPVPPRAHRTIRRPRLQRTAAKPARY	240
	orf132.pep	FGQRLLDAGGKIRHGRTRA	259
15	orf132ng	FGQRLLDAGGKIRHRLADW	261

An ORF132ng nucleotide sequence <SEQ ID 871> was predicted to encode a protein having amino acid sequence <SEQ ID 872>:

	1	MKHIHIIGIG	GTFMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TQLEALGIGV
20	51	HEGFDAQAQLE	EFQADIYVIG	NVARRGMDVV	EAILNRGLPY	ISGPQWLAEN
	101	VLHHHWLVGV	AGTHGKTTTA	SMLAWVLEYA	GLAPGFLIGG	VPKGFRFRFP
	151	PTANAASRPE	QQIAVFRHRS	RRIRHRLFRQ	TLQIRALSPA	YRRVEQSGIR
	201	PRRHLRLGR	DTDPVPPRA	HRTIRRPRL	QRTAAKPARY	FGQRLLDAGG
	251	KIRHRLAD	W*			

Further work revealed the following gonococcal DNA sequence <SEQ ID 873>:

25	1	ATGAAACACA	TCCACATTAT	CGGTATCGGC	GGCACGTTTA	TGGGCGGGAT
	51	TGCCGCCATT	GCCAAAGAAG	CCGGTTCAA	AGTCAGCGGT	TGCGACGCGA
	101	AGATGTATCC	GCCGATGAGC	ACCCAGCTCG	AAGCCTTGGG	CATAGGCGTA
	151	CACGAAGGCT	TCGATGCCGC	GCAGTGGAA	GAATTTCAAG	CCGATATTTA
	201	CGTCATCGGC	AATGTCGCCA	GGCGCGGGAT	GGATGTGGTC	GAGGCGATTT
30	251	TGAACCGTGG	GCTGCCTTAT	ATTTCCGGCC	CGCAATGGCT	GGCTGAAAac
	301	GTGctgcacc	atcaTTGGgt	ACTCGGCGTG	GcagggacGC	ACGgcaaAac
	351	gaccaCcGcg	tCCATGCTCG	CCTGGGTCTT	GGAATATGCC	GGACTCGCGC
	401	CGGGCTTCCT	CATCGGCGGT	gtaccggaAA	ATTTCCGGCGT	TTCCGCCCGC
	451	CTACCGCAAA	CGCCGCGTCA	AGACCCGAAC	AGCAAATCGC	CGTTTTTCGT
35	501	CATCGAAGCC	GACGAATACG	ACACCGCCTT	TTTCGACAAA	CGCTCCAAAT
	551	TCGTGCATTA	TCGCCCCGCT	ACCGCCGTGT	TGAACAATCT	GGAATTCGAC
	601	CACGCCGACA	TCTTCGCCGA	CTTGGGCGCG	ATACAGACCC	AGTTCACCA
	651	CCTCGTGCGC	ACCGTACCAT	CCGAAGGCCT	CATCGTCTGC	AACGGACAGC
	701	AGCAAAGCCT	GCAAGATACT	TTGGACAAAG	GCTGCTGGAC	GCCGGTGGAA
40	751	AAATTCGGCA	CCGGACACGG	CTGGCAGATT	GGTGAAGTCA	ATGCCGACGG
	801	CTCGTTCGAC	GTATTGCTTG	ACGGCAAAAA	AGCCGGACAC	GTCCGATGGG
	851	ATTTGATGGG	CGGACACAAC	CGCATGAACG	CGCTCGCCGT	CATCGCTGCC
	901	GCACGCCATG	CCGGAGTCGA	TGTTCCAGACG	GCCTGCGAAG	CCTTGGGTGC
	951	GTTTAAAAAC	GTCAAACGCC	GCATGGAAAT	CAAAGGCACG	GCAAACGGCA
45	1001	TCACCGTTTA	CGACGATTTC	GCCCACCACC	CGACCGCCAT	CGAAACCACG
	1051	ATTCAAGGTT	TGCGCCAACG	TGTCGGCGCG	GCGCGCATCC	TCGCCGTCTC
	1101	CGAGCCGCGT	TCCAACACCA	TGAAACTCGG	CACGATGAAG	TCCGCCCTGC
	1151	CCGCAAGCCT	CAAAGAAGCC	GACCAAGTGT	TCTGCTACGC	CGGCGGCGCG
	1201	GACTGGGACG	TTGCCGAAGC	CCTCGCGCCT	TTGGGCTGCA	GGCTGCGCGT
50	1251	CGGTAAAGAT	TTGATACCT	TCGTTGCCGA	AATTGTGAAA	AACGCCCGAA
	1301	CCGGCGACCA	TATTTTGGTG	ATGAGCAACG	GCGGTTTCGG	CGGAATACAC
	1351	ACCAAACCTGC	TGGACGCTTT	GAGATAG		

This corresponds to the amino acid sequence <SEQ ID 874; ORF132ng-1>:

	1	MKHIHIIGIG	GTFMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TQLEALGIGV
55	51	HEGFDAQAQLE	EFQADIYVIG	NVARRGMDVV	EAILNRGLPY	ISGPQWLAEN
	101	VLHHHWLVGV	AGTHGKTTTA	SMLAWVLEYA	GLAPGFLIGG	VPENFGVSAR
	151	LPQTTPRQDPN	SKSPFFVIEA	DEYDTAFFDK	RSKFVHYRPR	TAVLNNLEFD
	201	HADIFADLGA	IQTQFHHLVR	TVPSEGLIVC	NGQQQSLQDT	LDKGCWTFVE
	251	KFGTGHGWQI	GEVNADGSFD	VLLDGKKAGH	VAWDLMGHGN	RMNALAVIAA
60	301	ARHAGVDVQT	ACEALGAFKN	VKRRMEIKGT	ANGITVYDDF	AHHPTAIETT
	351	IQGLRQRVGG	ARILAVLEPR	SNTMKLGTMK	SALPASLKEA	DQVFCYAGGA
	401	DWDVAEALAP	LGCLRLVKGK	FDTFVAEIVK	NARTGDHILV	MSNGGFGGGIH
	451	TKLLDALR*				



ORF132ng-1 and ORF132-1 show 93.2% identity in 458 aa overlap:

```

5  orf132ng-1.pep MKHIHIIGIGGTFMGGIAAIAKEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAQAQLE
    orf132-1      MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD

10 orf132ng-1.pep EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAENVLHHHWVLGVAGTHGKTTTA
    orf132-1      EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA

15 orf132ng-1.pep SMLAWVLEYAGLAPGFLIGGVPENFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDK
    orf132-1      SMLAWVLEYAGLAPGFLIGGVPENFGVSARLPQTPRQDPNSQSPFFVIEADEYDTAFFDK

20 orf132ng-1.pep RSKFVHYRPTAVLNNLEFDHADIFADLGAIQTQFHHLVVRTVPSEGLIVCNGQQQSLQDT
    orf132-1      RSKFVHYRPTAVLNNLEFDHADIFADLGAIQTQFHYLVVRTVPSEGLIVCNGRQQSLQDT

25 orf132ng-1.pep LDKGCWTPVEKFGTGHGWQIGEVNADGSFDVLLDGKKAGHVAWDLMGHNRNMNALAVIAA
    orf132-1      LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKTAGRVKWDLMGRHNRNMNALAVIAA

30 orf132ng-1.pep ARHAGVDVQTACEALGAFKNVKKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
    orf132-1      ARHVGVDIQTACEALGAFKNVKKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG

35 orf132ng-1.pep ARILAVLEPRSNMTMKLGTMKXSALPASLKEADQVFCYAGGADWDVAEALAPLGCRRLRVGKD
    orf132-1      ARILAVLEPRSNMTMKLGTMKXSALPVSLKEADQVFCYAGGVDWDVAEALAPLGGRNLNVGKD

40 orf132ng-1.pep FDTFVAEIVKNARTGDHILVMSNGGFGGIHTKLLDALRX
    orf132-1      FDAFVAEIVKNAEVGDHILVMSNGGFGGIHGKLLLEALRX

```

In addition, ORF132ng-1 is homologous to a hypothetical *E.coli* protein:

```

35  pir||S56459 hypothetical protein o457 - Escherichia coli >gi|537075 (U14003)
    ORF o457 [Escherichia coli] >gi|1790680 (AE000494) hypothetical 48.5 kD protein
    in fbp-pmba intergenic region [Escherichia coli] Length = 457
    Score = 474 bits (1207), Expect = e-133
    Identities = 249/439 (56%), Positives = 294/439 (66%), Gaps = 13/439 (2%)

40  Query: 22  KEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAQAQLEEFQADIYVIGNVARRGMDVVE 81
    Sbjct: 21  RQLGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-QPDLVIIGNAMTRGNPCVE 79

45  Query: 82  AILNRGLPYISGPQWLAENVLHHHWVLGVAGTHGKTTASMLAWVLEYAGLAPGFLIGGV 141
    Sbjct: 80  AVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMATWILEQCQYKPGFVIGGV 139

50  Query: 142  PENFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDKRSKFVHYRPTAVLNNLEFDH 201
    Sbjct: 140  PGNFEVSAHL-----GESDFFVIEADEYDCAFFDKRSKFVHYCPRTLILNNLEFDH 190

55  Query: 202  ADIFADLGAIQTQFHHLVVRTVPSEGLIVCNGQQQSLQDTLDKGCWTPVEKFGTGHGWQIG 261
    Sbjct: 191  ADIFDDLKAIQKQFHHLVRIVPGQGRIIWPENDINLKQTMAMGCWSEQELVGEQGHWQAK 250

60  Query: 262  EVNADGS-FDVLLDGKKAGHVAWDLMGHNRNMNALAVIAAARHAGVDVQTACEALGAFKN 320
    Sbjct: 251  KLTTDASEWEVLLDGEKVGKWSLVEHNMHNLMAIAAARHVGVPADAANALGSFIN 310

65  Query: 321  VKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG-ARILAVLEPRSNMTMKLGTM 379
    Sbjct: 311  ARRLLELRGEANGVTVYDDFAHHPTAILATLAALRGKVGGTARIIAVLEPRSNMTMKMGIC 370

    Query: 380  KSALPASLKEADQVF-CYAGGADWDVAEALAPLGCRRLRVGKDFDTFVAEIVKNARTGDHI 438
    Sbjct: 371  KDDLAPSLGRADEVFLLQPAHIPWQVAEVAEACVQPAHWSGDVDTLADMVVKTAQPGDHI 430

    Query: 439  LVMSNGGFGGIHTKLLDAL 457

```

LVMSNGGFGGIH KLLD L  
 Sbjct: 431 LVMSNGGFGGIHQKLLDGL 449

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 ORF132-1 (26.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 20A shows the results of affinity purification of the His-fusion protein, and Figure 20B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 20C) and ELISA (positive result). These  
 10 experiments confirm that ORF132 is a surface-exposed protein, and that it is a useful immunogen.

### Example 103

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 875>

```

1  ..CCGGGCTATT ACGGCTCGGA TGACGAATTT AAGCGGGCAT TCGGAGAAAA
51  CTCGCCGACA TmCAAGAAAC ATTGCAACCG GAGCTGCGGG ATTTATGAAC
15  101  CCGTATTGAA AAAATACGGC AAAAAGCGCG CCAACAACCA TTCGGTCAGC
    151  ATTAGTGGCG ACTTCGGCGA TTATTTTCATG CCGTTCGCCA GCTATTCGCG
    201  CACACACCGT ATGCCCAACA TCCAAGAAAT GTATTTTTC CAAATCGGCG
    251  ACTCCGGCGT TCACACCGCC TTAACCAG AGCGCGCAA CACTTGCGAA
20  301  TTTGGCTTCr ATACCTATAA AAAAGGATTG TTAACAAG ATGATACATT
    351  AGGATTAAAA CTGGTCGGCT ACCGCGCCG CATCGACAAC TACATCCACA
    401  ACGTTTACGG GAAATGGTGG GATTTGAACG GGGATATTCC GAGCTGGGTC
    451  AGCAGCACCG GGCTTGCCTA CACCATCCAA CATCGCrATT TCaWAGACAA
    501  AGTGCATCAA nnnnnnnnnnn nnnnnnnnnnn nnnnnTACGAT TATGGGCGTT
25  551  TTTTCACCAA CCTTCTTAC GCCTATCAAA AAAGCACGCA ACCGACCAAC
    601  TTCAGCGATG CGAGCGAATC GCCCAACAAT GCGTCCAAAG AAGACCAACT
    651  CAAACAAGGT TATGGGTTGA GCAGGGTTTC CGCCCTGCCG CGAGATTACG
    701  GACGTTTGGG AGTCGGTACG CGCTGGTTGG GCAACAACT GACTTTGGGG
    751  GGCGCGATGC GCTATTTCCG CAAGAGCATC CGCGCGACGG CTGAAGAAGC
30  801  CTATATCGAC GGCACCAACG GGGGAAATAC CAGCAATTTC CGGCAACTGG
    851  GCAAGCGTTC CATCAAACAA ACCGAACTC TTGCCCGCCA GCCTTTGATT
    901  TTWgATTTTa ACGCCGCTTA CGAGCCGAAG AAAAACCTTA TTTTCCGCGC
    951  CGAAGTCAAA AATCTGTTCG ACAGGCGTTA TATCGATCCG CTCGATGCGG
1001 GCAATGATGC GGCAAC.GAG CGTTATTACA GCTCGTTTCA CCCGAAAGAC
1051 AAGGACrrAG AGCTAACGTG TAATGCTGAT AAAACGTTGT GCaACGGCAA
35  1101 ATACGGCGGC ACAAGCAAAA GCGTATTGAC CAATTTTGCA CGCGGACGCA
    1151 CCTTTTgAT GACGATGAGC TACAAGTTTT AA
  
```

This corresponds to the amino acid sequence <SEQ ID 876; ORF133>:

```

1  ..PGYYGSDDEF KRAFGENSPT XKKHCNRSCG IYEPVLKKYG KKRANNHSVS
51  ISADFGDYFM PFASYSRTHR MPNIQEMYFS QIGDSGVHTA LKPERANTWQ
40  101  FGFXTYKKGL LKQDPTLGLK LVGYRSRIDN YIHNVYGKWW DLNGDIPSWV
    151  SSTGLAYTIQ HRXFXDKVHQ XXXXXXXXX YGRFFTNLSY AYQKSTQPTN
    201  FSDASESPNN ASKEDQLKQG YGLSRVSALP RDYGRLEVGT RWLGNKLTIG
    251  GAMRYFGKSI RATAERYID GTNGGNTSNF RQLGKRSIKQ TETLARQPLI
    301  XDFNAAYEPK KNLFRAEVK NLFDRRYIDP LDAGNDAAE RYYSSFDPKD
45  351  KDXDVTGNAD KTLNCKYGG TSKSVLTNFA RGRTFLMTMS YKF*
  
```

Further work revealed the further partial DNA sequence <SEQ ID 877>:

```

1  GAGGCGCAGA TACAGGTTTT GGAAGATGTG CACGTCAAGG CGAAGCGCGT
51  ACCGAAAGAC AAAAAAGTGT TTACCGATGC GCGTGCCGTA TCGACCCGTC
50  101  AGGATATATT CAAATCCAGC GAAAACCTCG ACAACATCGT ACGCAGCATC
    151  CCCGGTGCGT TTACACAGCA AGATAAAAGC TCGGGCATTG TGTCTTTGAA
    201  TATTCGCGGC GACAGCGGGT TCGGGCGGGT CAATACGATG GTGACGGCA
    251  TCACGCAGAC CTTTTATTTCG ACTTCTACCG ATGCGGGCAG GGCAGGCGGT
  
```

	301	TCATCTCAAT	TCGGTGCATC	TGTCGACAGC	AATTTTATTG	CCGGACTGGA
	351	TGTCGTCAAA	GGCAGCTTCA	GCGGCTCGGC	AGGCATCAAC	AGCCTTGCCG
	401	GTTCGGCGAA	TCTGCGGACT	TTAGGCGTGG	ATGACGTCGT	TCAGGGCAAT
5	451	AATACCTACG	GCCTGCTGCT	AAAAGGTCTG	ACCGGCACCA	ATTCAACCAA
	501	AGGTAATGCG	ATGGCGGCGA	TAGGTGCGCG	CAAATGGCTG	GAAAGCGGAG
	551	CATCTGTCCG	TGTGCTTTAC	GGGCACAGCA	GGCGCAGCGT	GGCGCAAAAT
	601	TACCGCGTGG	GCGGCGGCGG	GCAGCACATC	GGAAATTTTG	GCGCGGAATA
	651	TTTGAACCGG	CGCAAGCAGC	GATATTTTGT	ACAAGAGGGT	GCTTTGAAAT
	701	TCAATTCCGA	CAGCGGAAAA	TGGGAGCGGG	ATTTACAAAG	GCAACAGTGG
10	751	AAATACAAGC	CGTATAAAAA	TTACAACAAC	CAAGAACTAC	AaAAATACAT
	801	CGAAGAGCAT	GACAAAAGCT	GGCGGGAAAA	CCTg . CaCCG	CAATACGACA
	851	TTACCCCAT	CGATCCGTCC	AGCCTGAAGC	AGCAGTCGGC	AGGCAATCTG
	901	TTTAAATTGG	AATACGACGG	CGTATTTCAAT	AAATACACGG	CGCAATTTTCG
	951	CGATTTAAAC	ACCAAAATCG	GCAGCCGCAA	AATCATCAAC	CGCAATTATC
15	1001	AGTTCAATTA	CGGTTTGTCT	TTGAACCCGT	ATACCAACCT	CAATCTGACC
	1051	GCAGCCTACA	ATTGCGGCAG	GCAGAAATAT	CCGAAAGGGT	CGAAGTTTAC
	1101	AGGCTGGGGG	CTTTTAAAGG	ATTTTGAAAC	CTACAACAAC	GCGAAAATCC
	1151	TCGACCTCAA	CAACACCGCC	ACCTTCCGGC	TGCCCCGCGA	AACCGAGTTG
	1201	CAAACCACTT	TGGGCTTCAA	TTATTTCCAC	AACGAATACG	GCAAAAACCG
20	1251	CTTTCTGAA	GAATTGGGGC	TGTTTTTCGA	CGGTCCTGAT	CAGGACAACG
	1301	GGCTTTATTC	CTATTTGGGG	CGGTTTAAGG	GCGATAAAGG	GCTGCTGCCC
	1351	CAAAAATCAA	CCATTGTCCA	ACCGGCCGGC	AGCCAATATT	TCAACACGTT
	1401	CTACTTCGAT	GCCGCGCTCA	AAAAAGACAT	TTACCGCTTA	AACTACAGCA
	1451	CCAATACCGT	CGGCTACCGT	TTCGGCGGCG	AATATACGGG	CTATTACGGC
25	1501	TCGGATGACG	AATTAAAGCG	GGCATTTCGA	GAAAACTCGC	CGACATACAA
	1551	GAAACATTGC	AACCGGAGCT	GCGGGATTTA	TGAACCCGTA	TTGAAAAAAT
	1601	ACGGCAAAAA	GCGCGCCAAC	AACCATTTCG	TCAGCATTAG	TGCGGACTTC
	1651	GGCGATTATT	TCATGCCGTT	CGCCAGCTAT	TCGCGCACAC	ACCGTATGCC
	1701	CAACATCCAA	GAAATGTATT	TTTCCCAAT	CGGCGACTCC	GGCGTTTACA
30	1751	CCGCCTTAAA	ACCAGAGCGC	GCAAAACACTT	GGCAATTGG	CTTCAATTACC
	1801	TATAAAAAAG	GATTGTATAA	ACAAGATGAT	ACATTAGGAT	TAAAACCTGGT
	1851	CGGCTACCGC	AGCCGCATCG	ACAACTACAT	CCACAACGTT	TACGGGAAAT
	1901	GGTGGGATTT	GAACGGGGAT	ATTCCGAGCT	GGGTACGAG	CACCGGGCTT
	1951	GCCTACACCA	TCCAACATCG	CAATTTCAAA	GACAAAGTGC	ACAAACACGG
35	2001	TTTTGAGTTG	GAGCTGAATT	ACGATTATGG	GCGTTTTTTC	ACCAACCTTT
	2051	CTTACGCCTA	TCAAAAAAGC	ACGCAACCGA	CCAACCTCAG	CGATGCGAGC
	2101	GAATCGCCCA	ACAATGCGTC	CAAAGAAGAC	CAACTCAAAC	AAGGTTATGG
	2151	GTTGAGCAGG	GTTTCCGCCC	TGCCGCGAGA	TTACGGACGT	TTGGAAGTCG
40	2201	GTACGCGCTG	GTTGGGCAAC	AAACTGACTT	TGGGCGGCGC	GATGCGCTAT
	2251	TTCCGCAAGA	GACGCTGAA	GAACGCTATA	TGACGCGCAC	TGACGCGCAC
	2301	CAACGGGGGA	AATACCAGCA	ATTTCCGGCA	ACTGGGCAAG	CGTTCCATCA
	2351	AACAAACCGA	AACTCTTGCC	CGCCAGCCTT	TGATTTTTGA	TTTTTACGCC
	2401	GCTTACGAGC	CGAAGAAAAA	CCTTATTTTC	CGCGCCGAAG	TCAAAAATCT
	2451	GTTTCGACAGG	CGTTATATCG	ATCCGCTCGA	TGCGGGCAAT	GATGCGGCAA
45	2501	CGCAGCGTTA	TTACAGCTCG	TTCGACCCGA	AAGACAAGGA	CGAAGACGTA
	2551	ACGTGTAATG	CTGATAAAAC	GTTGTGCAAC	GGCAAATACG	GCGGCACAAG
	2601	CAAAAGCGTA	TTGACCAATT	TTGCACGCGG	ACGCACCTTT	TTGATGACGA
	2651	TGAGCTACAA	GTTTAA			

This corresponds to the amino acid sequence <SEQ ID 878; ORF133-1>:

50	1	EAQIQVLEDV	HVKAKRVPKD	KKVFTDARAV	STRQDIFKSS	ENLDNIVRSI
	51	PGAFTQQDKS	SGIVSLNIRG	DSGFGRVNTM	VDGITQTFYS	TSTDAGRAGG
	101	SSQFGASVDS	NFIAGLDVVK	GSFSGSAGIN	SLAGSANLRT	LGVDDVVQGN
	151	NTYGLLLKGL	TGTNSTKGNA	MAAIGARKWL	ESGASVGVLY	GHSRRSVAQN
55	201	YRVGGGGQHI	GNFGAEYLER	RKQRYFVQEG	ALKFNSDSGK	WERDLQRQQW
	251	KYKPYKNYNN	QELQKYIEEH	DKSWRENLXP	QYDITPIDPS	SLKQQSAGNL
	301	FKLEYDGVFN	KYTAQFRDLN	TKIGSRKIIN	RNYQFNYGLS	LNPTYNLNLT
	351	AAYNSSGRQKY	PKGSKFTGWG	LLKDFETYN	AKILDNLNNTA	TFRLPRETEL
	401	QTTLGFNYFH	NEYGKNRFPE	ELGLFFDGP	QDNGLYSYLG	RFGDKGLLP
60	451	QKSTIVQPAG	SOYFNTFFFD	AALKKDIYRL	NYSTNTVGYR	FGGEYTYGYG
	501	SDDEFKRAFG	ENSPYKKHC	NRSCGIYEPV	LKKYGKKRAN	NHSVVISADF
	551	GDYFMPFASY	SRTHRPNQI	EMYFSQIGDS	GVHTALKPER	ANTWQFGFNT
	601	YKKGLLKQDD	TLGLKLVGYS	SRIDNYIHNV	YKQWDLNGD	IPSWVSSTGL
	651	AYTIQHRNFK	DKVHKHGFEL	ELNYDYGRFF	TNLSYAYQKS	TQPTNFSNAS
	701	ESPNNASKED	QLKQGYGLSR	VSALPRDYGR	LEVGTWRLGN	KLTLLGAMRY
65	751	FGKSIRATAE	ERYIDGTNGG	NTSNFRQLGK	RSIKQTETLA	RQPLIFDFYA
	801	AYEPKKNLIF	RAEVKNLFDR	RYIDPLDAGN	DAATQRYYS	FDPKDKDEDV
	851	TCNADKTLN	GKYGTSKSV	LTNFARGRTF	LMTMSYKF*	

Computer analysis of this amino acid sequence gave the following results:

Homology with the probable TonB-dependent receptor HI121 of *H. influenzae* (accession number U32801)

ORF133 and HI121 show 57% aa identity in 363aa overlap:

```

5      Orf133: 31  IYEPVLKKYKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTA 90
        I EP+L K G K+A NHS ++SA+ DYFMPF +YSRTHRMPNIQEM+FSQ+ ++GV+TA
      HI121: 563  INEPILHKS GHKKA FNHSATLSAELSDYFMPFFTYSRTHRMPNIQEMFFSQVSNAGVNTA 622

10     Orf133: 91  LKPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHN VYGKWWDLNGDIPSWV 150
        LKPE+++T+Q GF TYKKGL QDD LG+KLVGYRS I NYIHN VYG WW +P+W
      HI121: 623  LKPEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFIKNYIHN VYG VWW--RDGMPTWA 680

15     Orf133: 151  SSTGLAYTIQHRXFXDKVHXXXXXXXXXYDYGRFFTNLSYAYQKSTQPTNFS DASESPNN 210
        S G YTI H+ + V YD GRFF N+SYAYQ++ QPTN++DAS PNN
      HI121: 681  ESNQFKYTIHQNYKPIVKKSGVELEIN YDMGRFFANVSYAYQRTNQPTNYADASPRPNN 740

20     Orf133: 211  ASKEDQLKQGYGLSRVSALPRDYGRLEVGT RWLGNKLT LGGAMRYFGKSIRATAEERYID 270
        AS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLT LG A RY+GKS RAT EE YI+
      HI121: 741  ASQEDILKQGYGLSRVSMLPKDYGRLEGT RWFDQKLT LGLAARYYGKSKRATIEEYIN 800

25     Orf133: 271  GTNGGNTSNFRQLGKRSIKOTETLARQPLIXDFNAA YEPKKNLIFRAEVKNLFDRRYIDP 330
        G+ + R+ ++K+TE + +QP+I D + +YEP K+LI +AEV+NL D+RY+DP
      HI121: 801  GSR-FKKNL LRRENYAVKKTEDIKKQPIILDHVS YEPIKDLIIKAEVQNLLDKRYVDP 859

30     Orf133: 331  LDAGNDAAXERYYSFDPKDKDXDVT CNADKTL CNKGKYGTSKSVLTNFARGRTFLMTMS 390
        LDAGNDAA +RYYSS + + C D + C GG+ K+VL NFARGRT+++++
      HI121: 860  LDAGNDAASQRYSS L-----NNSIECAQDSSAC-----GGSDKTVLYNFARGRTYILSLN 910

      Orf133: 391  YKF 393
        YKF
      HI121: 911  YKF 913

```

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF133 shows 90.8% identity over a 392aa overlap with an ORF (ORF133a) from strain A of *N.*

*meningitidis*:

```

35     orf133.pep          10      20      30
        PGYYGSDDEFKRAFGENSPTXKKHCNRCGI
      orf133a          FYFDAALKKDIYRLNYSTNTVGYRFGGX YTGYYXSDDEFKRAFGENSPTXKKHCNQSCGI
        450      460      470      480      490      500

40     orf133.pep          40      50      60      70      80      90
        YEPVLKKYKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL
      orf133a          YEPVLKKYKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL
        510      520      530      540      550      560

45     orf133.pep          100     110     120     130     140     150
        KPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHN VYGKWWDLNGDIPSWVS
      orf133a          KPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDXYIHN VYGKWWDLNGNIPSWVS
        570      580      590      600      610      620

50     orf133.pep          160     170     180     190     200     210
        STGLAYTIQHRXFXDKVHXXXXXXXXXYDYGRFFTNLSYAYQKSTQPTNFS DASESPNNA
      orf133a          STGLAYTIQHRNFKDKVHKHGFEELELN DYXRFFTNLSYAYQKSTQPTNFS DASESPNNA
        630      640      650      660      670      680

55     orf133.pep          220     230     240     250     260     270
        SKEDQLKQGYGLSRVSALPRDYGRLEVGT RWLGNKLT LGGAMRYFGKSIRATAEERYIDG
      orf133a          SKEDQLKQGYGLSRVSALPRDYGRLEVGT RWLGNKLT LGGAMRYFGKSIRATAEERYIDX
        690      700      710      720      730      740

60     orf133.pep          280     290     300     310     320     330
        TNGGNTSNFRQLGKRSIKOTETLARQPLIXDFNAA YEPKKNLIFRAEVKNLFDRRYIDPL
      orf133a          TNGGNTSNFRQLGKRSIKOTETLARQPLIXDFNAA YEPKKNLIFRAEVKNLFDRRYIDPL

```

[illegible]

A partial ORF133a nucleotide sequence <SEQ ID 879> is:

	1	AAAGACAAAA	AAGTGTTTAC	CGATGCGCGT	GCCGTATCGA	CCCGTCAGGA
	51	TATATTCAAA	TCCANCGAAA	ACCTCGACAA	CATCGTACGC	ANCATCCCCG
20	101	GTGCGTTTAC	ACANCAANAT	AAAAGCTCGG	GCNTTGTGTC	TTTGAATATT
	151	CGCNGCGACA	GCGGGTTCCG	GCGGGTCAAT	ACNATGGTNG	ACGGCATCAC
	201	NCANACCTT	TATTCGACTT	CTACCGATGC	GGGCAGGGCA	GCGGGTTTAT
	251	CTCAATTCCG	TGCATCTGTC	GACAGCAATT	TTATNGCCGA	ACTGGATGTC
	301	GTCAAAGGCA	GCTTCAGCGG	CTCGGCAGGC	ATCAACAGCC	TTGCCGGTTC
25	351	GGCGAATCTG	CGGACTTTAN	GCGTGGATGA	TGTCGTTTAC	GGCAATANTA
	401	CNTACGGCCT	GCTGCTAAAA	GGTCTGACCG	GCACCAATTG	AACCAAAAGG
	451	AATGCGATGG	CGGCGATAGG	TGCGCGCAAA	TGGCTGGAAA	GCGGAGCATC
	501	TGTCGGGTGTG	CTTTACGGGC	ACAGCAGGCG	CAGCGTGGCG	CAAAATTACC
	551	GCGTGGGCGG	CGGCGGGCAG	CACATCGGAA	ATTTTGGCGC	GGAATATCTG
30	601	GAACGACGCA	AGCAACGATA	TTTTTGAGCA	GAAGGCGGGT	TGAAATTCAA
	651	TTCCAACAGC	GGAAAAATGG	AGCGGGATT	CCAAAAGTCG	TACTGGAAAA
	701	CCAAGTGGTA	TCAAAAAATC	GATGCCCCCC	AGAAGTATGCA	AAAATACATC
	751	GAAGGTCATG	ATAAAAGCTG	GCGGGAAC	CTGGCGCCGC	AATACGACAT
	801	CACCCCATC	GATCCGTC	GCCTGAAGCN	GCAGTCGGCA	GGCAACCTGT
35	851	TTAAATTGGA	ATACGACGGC	GTATTCAATA	AATACACGGC	GCAATTTTCG
	901	GATTTAAACA	CGAAAAATCG	CAGCCGCAAA	ATCATCAACC	GCAATTTATCA
	951	ATTCAATTAC	GTTTGTCTT	TGAACCCGTA	TACCAACCTC	ACTCTGACCG
	1001	CAGCCTACAA	TTCGGGCAGG	CAGAAATATC	CGAAAGGGTC	GAAGTTTACA
	1051	GGCTGGGGGC	TTTTNAAAGA	TTTTGAAACC	TACACAACG	CAAAAATCCT
40	1101	CGACCTCANC	AACACCTCCA	CCTTCCGGCT	GCCCCGTGAA	ACCGAGTTGC
	1151	AAACCACTT	GGGCTTCAAT	TATTTCCACA	ACGAATACGG	CAAAAACCGC
	1201	TTTCTGAAG	AATTGGGGCT	GTTTTTCGAC	GGTCCGGATC	ANGACAACGG
	1251	GCTTTATTCC	TATTTGGGGC	GGTTTAAGGG	CGATAAAGGG	CTGCTGCCCC
	1301	AAAAATCAAC	CATTGTCCAA	CCGGCCGGCA	GCCAATATTT	CAACACGTTT
45	1351	TACTTCGATG	CCGCGCTCAA	AAAAGACATT	TACCGCTTAA	ACTACAGCAC
	1401	CAATACCGTG	GGCTACCGTT	TCGGCGGCNA	ATATACGGG	TATTACNGCT
	1451	CGGATGACGA	ATTTAAGCGG	GCATTCCGGAG	AAAACCTCGC	GACATACANG
	1501	AAACATTGCA	ACCAGAGCTG	CGGAATTTAT	GAACCCGTAT	TGAAAAAATA
	1551	CGGCAAAAAG	CGCGCCAAAC	ACCATTTCGGT	CAGCATTAGT	GCCGACTTTCG
50	1601	GCGATTATTT	CATGCCGTTT	GCCAGCTATT	CGCGCACACA	CCGTATGCC
	1651	AACATCCAA	AAATGTATTT	TTCCCAATC	GGCGACTCCG	CCGTTTACAC
	1701	CGCCTTAAAA	CCAGAGCGCG	CAAACACTTG	GCAATTTGGC	TTCAATACCT
	1751	ATAAAAAAGG	ATTGTTAAAA	CAAGATGATA	TATTAGGATT	AAAACCTGGT
	1801	GGTACCGCA	GCCGCATCGA	CNACTACATC	CACAACGTTT	ACGGGAAATG
55	1851	GTGGGATTTG	AACGGGAATA	TTCCGAGCTG	GGTCAGCAGT	ACCGGGCTTG
	1901	CCTACACCAT	CCAACACCGC	AATTTCAAAG	ACAAAGTGCA	CAAACACGGT
	1951	TTTGAGTTGG	AGCTGAATTA	CGATTATNGG	CGTTTTTTTCA	CCAACCTTTC
	2001	TTACGCCTAT	CAAAAAAGCA	CGCAACCGAC	CAACTTCAGC	GATGCGAGCG
	2051	AATCGCCCAA	CAATGCGTCC	AAAGAAGACC	AACTCAAACA	AGGTTATGGG
60	2101	TTGAGCAGGG	TTTCCGCCCT	GCCGCGAGAT	TACGGACTTT	TGGAGTCCGG
	2151	TACGCGCTGG	TTGGGCAACA	AACTGACTTT	GGGCGCGCGG	ATGCGCTATT
	2201	TCGGCAAGAG	CATCCGCGCG	ACGGCTGAAG	AACGCTATAT	CGACGNCACC
	2251	AATGGGGNAN	NTACCAGCAA	TTTCCGGCAA	CTGGGCAAGC	GTTCATCAN
	2301	ACAAACCGAA	ACCTTTGCCC	GCCAGCCTTT	GATTTTGTAT	TTNTACCGCC
65	2351	CTTACGAGCC	GAGAAAAAAN	CTTATTTTCC	GCCGCGGAAGT	CAAAAATCTG
	2401	TTGCACAGGC	GTTATATCGA	TCCGCTCGAT	GCGGGCAATG	ATGCGGCAAC
	2451	GCAGCGTTAT	TACAGTTCGT	TCGACCCGAA	AGACAAGGAC	GAAGAAGTAA
	2501	CGTGTAATGA	TGATAACACG	TTATGCAACG	GCAAATACGG	CGGCACAAGC
	2551	AAAAGCGTAT	TGACCAATTT	TGCACGCGGA	CNCACCTTTT	TGATAACGAT
	2601	GAGCTACAAG	TTTTTAA			

This encodes a protein having (partial) amino acid sequence <SEQ ID 880>:

```

      1 KDKKVFTDAR AVSTRQDIFK SXENLDNIVR XIPGAFTXQX KSSGXVSLNI
     51 RXDSGFGRVN TMVDGITXTF YSTSTDAGRA GGSSQFGASV DSNFXAGLDV
    101 VKGSFSGSAG INSLAGSANL RTLXVDDVVQ GNXTYGLLLK GLTGTNSTKG
5     151 NAMAAGARK WLESASVGV LYGHSRRSVA QNYRVGGGGQ HIGNFGAEYL
    201 ERRKQRYFEQ EGGLKFNSNS GKWERDFQKS YWKTWKYQKY DAPQELQKYI
    251 EGHDKSWREN LAPQYDITPI DPSSLKXQSA GNLFKLEYDG VENKYTAQFR
    301 DLNTKIGSRK IINRNYQFNY GLSLNPYTNL NLTAAYNSGR QKYPKGSKFT
    351 GWGLXKDFET YNNAKILDIX NTSTFRLPRE TELQTTLGFN YFHNEYGKNR
   10  401 FPEELGLFFD GPDNDGLYS YLGRFGDKG LLPQKSTIVQ PAGSQYFNTF
    451 YFDAALKKDI YRLNYSTNTV GYRFGGXYTG YYXSDDEFKR AFGENSPTYX
    501 KHCNQSCGIY EPVLKKGK RANNHSVSIS ADFGDYFMPF ASYSRTHRMP
    551 NIQEMYFSQI GDSGVHTALK PERANTWQFG FNTYKKGLLK QDDILGLKLV
    601 GYRSRIDXYI HNVYKWWDL NGNIPSWVSS TGLAYTIQHR NFKDKVHKHG
   15  651 FELELNIDYX RFFTNLSYAY QKSTQPTNFS DASESPNNAS KEDQLKQGYG
    701 LSRVSALPRD YGRLEVGTWR LGNKLTLGGA MRYFGKSIRA TAEERYIDXT
    751 NGXXTSNFRQ LGKRSIXQTE TLAQPLIFD XYAAYEPKKX LIFRAEVKNL
    801 FDRRYIDPLD AGNDAATQRY YSSFDPKDKD EEVTCNDNT LCNGKYGGTS
    851 KSVLTNFARG XTFLITMSYK F*

```

20 ORF133a and ORF133-1 show 94.3% identity in 871 aa overlap:

```

                                10      20      30      40
or133a.pep                    KDKKVFTDARAVSTRQDIFKSXENLDNIVRXIPGAFTXQXKS
                                |||||
25  or133-1                    EAQIQVLEDVHVKAARVPKDKKVFTDARAVSTRQDIFKSSSENLDNIVRSIPGAFTQQDKS
                                10      20      30      40      50      60

                                50      60      70      80      90      100
or133a.pep                    SGXVSLNIRXDSGFGRVNTMVDGITXTFYSTSTDAGRAGSSQFGASVDSNFXAGLDVVK
                                || |||||
30  or133-1                    SGIVSLNIRGDSGFGRVNTMVDGITQTFYSTSTDAGRAGSSQFGASVDSNFIAGLDVVK
                                70      80      90      100      110      120

                                110      120      130      140      150      160
or133a.pep                    GSFSGSAGINSLAGSANLRTLXVDDVVQGNXTYGLLLKGLTGTNSTKGNAMAAGARKWL
                                |||||
35  or133-1                    GSFSGSAGINSLAGSANLRTLGVDDVVQGNNTYGLLLKGLTGTNSTKGNAMAAGARKWL
                                130      140      150      160      170      180

                                170      180      190      200      210      220
40  or133a.pep                    ESGASVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFEQEGGLKFNSNSGK
                                |||||
or133-1                    ESGASVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFVQEGALKFNSDSGK
                                190      200      210      220      230      240

                                230      240      250      260      270      280
45  or133a.pep                    WERDFQKSYWKTWKYQKYDAPQELQKYIEGHDKSWRENLAQYDITPIDPSSLKXQSAGN
                                |||||
or133-1                    WERDLQRQWKYKPYKNYNN-QELQKYIEGHDKSWRENLAQYDITPIDPSSLKQQSAGN
                                250      260      270      280      290

                                290      300      310      320      330      340
50  or133a.pep                    LFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNYGLSLNPYTNLNLTAAYNSGRQK
                                |||||
or133-1                    LFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNYGLSLNPYTNLNLTAAYNSGRQK
55  300      310      320      330      340      350

                                350      360      370      380      390      400
or133a.pep                    YPKGSKFTGWGLXKDFETYNNAKILDIXNTSTFRLPRETELQTTLGFNHYFHNEYGKNRFP
                                |||||
60  or133-1                    YPKGSKFTGWGLLKDFETYNNAKILDIXNTATFRLPRETELQTTLGFNHYFHNEYGKNRFP
                                360      370      380      390      400      410

                                410      420      430      440      450      460
or133a.pep                    EELGLFFDGPDXNDGLYSYLGRFGDKGGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR
                                |||||
65  or133-1                    EELGLFFDGPDXNDGLYSYLGRFGDKGGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR
                                420      430      440      450      460      470

```

-478-

		470	480	490	500	510	520
	orf133a.pep	LNYSTNTVGYRFGGXYTGYYXSDDEFKRAFGENSPTYXKHCNQSCGIYEPVLKKYGKKRA					
	orf133-1	LNYSTNTVGYRFGGEYTGYYGSDDEFKRAFGENSPTYXKHCNRSCGIYEPVLKKYGKKRA					
5		480	490	500	510	520	530
		530	540	550	560	570	580
	orf133a.pep	NNHSVSIADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTALKPERANTWQFGFN					
10	orf133-1	NNHSVSIADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTALKPERANTWQFGFN					
		540	550	560	570	580	590
		590	600	610	620	630	640
	orf133a.pep	TYKKGLLKQDDILGLKLVGYRSRIDXYIHNVYGKWWDLNGNIPSWVSSTGLAYTIQHRNF					
15	orf133-1	TYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVSSTGLAYTIQHRNF					
		600	610	620	630	640	650
		650	660	670	680	690	700
20	orf133a.pep	KDKVHKHGFELNLDYXRFFTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQGYGLS					
	orf133-1	KDKVHKHGFELNLDYGRFFTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQGYGLS					
		660	670	680	690	700	710
25		710	720	730	740	750	760
	orf133a.pep	RVSALPRDYGRLEVGTWLGKLTGGAMRYFGKSIRATAEERYIDXTNGXXTSNFRQLG					
	orf133-1	RVSALPRDYGRLEVGTWLGKLTGGAMRYFGKSIRATAEERYIDGTNGGNTSNFRQLG					
30		720	730	740	750	760	770
		770	780	790	800	810	820
	orf133a.pep	KRSIXQTETLARQPLIFDXYAAYEPKKXLI FRAEVKNLFDRRYIDPLDAGNDAATQRYYS					
35	orf133-1	KRSIXQTETLARQPLIFDYAAYEPKKNLI FRAEVKNLFDRRYIDPLDAGNDAATQRYYS					
		780	790	800	810	820	830
		830	840	850	860	870	
	orf133a.pep	SFDPKDKDEEVTCDNDNTLCNGKYGGTSKSVLTNFARGXTFLITMSYKFX					
40	orf133-1	SFDPKDKDEDVTCNADKTLNCGKYGGTSKSVLTNFARGRTFLMTMSYKFX					
		840	850	860	870	880	

Homology with a predicted ORF from *N.gonorrhoeae*ORF133 shows 92.3% identity over 392 aa overlap with a predicted ORF (ORF133ng) from *N.*45 *gonorrhoeae*:

	orf133.pep	PGYYGSDDEFKRAFGENSPTYXKHCNRSCGI	31
	orf133ng	FYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKRAFGENSPAYKEHCDPSCGL	560
50	orf133.pep	YEPVLKKYGKKRANNHSVSIADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL	91
	orf133ng	YEPVLKKYGKKRANNHSVSIADFGDYFMPFAGYSRTHRMPNIQEMYFSQIGDSGVHTAL	620
55	orf133.pep	KPERANTWQFGFTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVS	151
	orf133ng	KPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVG	680
	orf133.pep	STGLAYTIQHRFXDKVHQXXXXXXXXDYGRFFTNLSYAYQKSTQPTNFSDAESPNNNA	211
60	orf133ng	STGLAYTIRHRNFKDKVHKHGFELNLDYGRFFTNLSYAYQKSTQPTNFSDAESPNNNA	740
	orf133.pep	SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGAMRYFGKSIRATAEERYIDG	271
	orf133ng	SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGAMRYFGKSIRATAEERYIDG	800
65	orf133.pep	TNGGNTSNFRQLGKRSIXQTETLARQPLIXDFNAAYEPKKNLI FRAEVKNLFDRRYIDPL	331
	orf133ng	TNGGNTSNVRQLGKRSIXQTETLARQPLIFDYAAYEPKKNLI FRAEVKNLFDRRYIDPL	860

```

orf133.pep  DAGNDAAXERYSSFDPKDKDXDVTCTNADKTLGNGKYGGTSKSVLTNFARGRTFLMTMSY 391
|||||::|:||||| ||||||| ||||||| ||||||| ||||||| |||||||
orf133ng    DAGNDAATQRYSSFDPKDKDEDVTCTNADKTLGNGKYGGTSKSVLTNFARGRTFLMTMSY 920
5
orf133.pep  KF 393
||
orf133ng    KF 922

```

The complete length ORF133ng nucleotide sequence <SEQ ID 881> is predicted to encode a protein having amino acid sequence <SEQ ID 882>:

```

1  MRSSFRLKPI CFYLMGVMLY HHSYAEDAGR AGSEAIQIVL EDVHVKAQRV
51  PKDKKVFTDA RAVSTRQDVF KSGENLDNIV RSIPGAFTQQ DKSSGIVSLN
101 IRGDSGFGRV NTMVDGITQT FYSTSTDAGR AGSSSQFGAS VDSNFIAGLD
151 VVKGSFSGSA GINSLAGSAN LRTLGVDDVV QGNNTYGLLL KGLTGTNSTK
201 GNAMAAIGAR KWLESGASVG VLYGHSRRGV AQNYRVGGGG QHIGNFGEEY
251 LERRKQYFV QEGGLKFNAG SGKWERDLQR QYWKTKWYKK YEDPQELQKY
301 IEEHDKSWRE NLAPQYDITP IDPSGLKQOS AGNLLNLEYD GFVNKYTAQF
351 RDLNTRIGSR KIINRNYQFN YGLSLNPYTN LNLTAAYNSG RQKYPKGAKF
401 TGWGLLKDFE TYNNAKILD NNTATFRLPR ETELQTTLGF NYFHNEYGKN
20 451 RFPEELGLFF DGPDQDNGLY SYLGRFKGDK GLLPQKSTIV QPAGSQYFNT
501 FYFDAALKKD IYRLNYSTNA INYRFGGEYT GYGSSENEFK RAFGENSPAY
551 KEHCDPSCGL YEPVLKKYKG KRANNHSVSI SADFGDYFMP FAGYSRTHRM
601 PNIQEMYFSQ IGDSGVHTAL KPERANTWQF GFNTYKKGLL KQDDILGLKL
651 VGYRSRIDNY IHNVYGKWD LNGDIPSWVG STGLAYTIRH RNFKDKVHKH
25 701 GFELELNIDY GRFFTNLSYA YQKSTQPTNF SDASESPNNA SKEDQLKQGY
751 GLSRVSALPR DYGRLEVGTW WLGKLTLLGG AMRYFGKSIR ATAEERYIDG
801 TNGGNTSNVR QLGKRSIKOT ETLARQPLIF DFYAAYPEPK NLIFRAEVKN
851 LFDRIYIDPL DAGNDAATQR YYSSFDPKDK DEDVTCNADK TLGNGKYGGT
901 SKSVLTNFAR GRTFLMTMSY KF*

```

A variant was also identified, being encoded by the gonococcal DNA sequence <SEQ ID 883>:

```

1  ATGAGATCTT CTTTCCGGTT GAAGCCGATT TGTTTTTATC TTATGGGTGT
51  TATGCTATAT CATCATAGTT ATGCCGAAGA TGCAGGGCGC GCGGGCAGCG
101 AGGCGCAGAT ACAGGTTTGT GAAGATGTGC ACGTCAAGGC GAAGCGCGTA
151 CCGAAAGACA AAAAAGTGTT TACCGATGCG CGTGCCGTAT CGACCCGTca
201 gGATGTGTTC AAATCCGGCG AAAACCTCGA CAACATCGTA CGCAGCATAC
251 CCGGTGCGTT TACACAGCAA GATAAAGCT CGGGCATTGT GTCTTTGAAT
301 ATTCGCGGCG ACAGCGGGTT CGGGCGGGTC AATACGATGG TGGACGGCAT
351 CAGCGAGACC TTTTATTCGA CTTCTACCGA TGCGGGCAGG GCAGCGGGTT
401 CATCTCAATT CCGTGCATCT GTCGACAGCA ATTTTATTGC CGGACTGGAT
451 GTCGTCAAAG GCAGCTTCAG CGGCTCGGCA GGCATCAACA GCCTTGCCGG
501 TTCGGCGAAT CTGCGGACTT TAGGCGTGGA TGACGTCGTT CAGGGCAATA
551 ATACCTACGG CCTGCTGCTA AAAGGTCTGA CCGGCACCAA TTCAACCAAA
601 GTCAATGCGA TGGCGGCGAT AGGTGCGCGC AAATGGCTGG AAAGCGGAGC
651 GTCTGTGGT GTGCTTTACG GGCACAGCAG GCGCGGCGTG GCGCAAAATT
45 701 ACCGCGTGGG CGGCGGCGGG CAGCACATCG GAAATTTTGG TGAAGAATAT
751 CTGGAACGGC GCAAACAGCA ATATTTTGTA CAAGAGGGTG GTTTGAAATT
801 CAATGCCGGC AGCGGAAAAT GGAACGGGA TTTGCAAAGG CAATACTGGA
851 AAACAAAGTG GTATAAAAAA TACGAAGACC CCAAGAAGT GCAAAAATAC
901 ATCGAAGAGC ATGATAAAG CTGGCGGGAA AACCTGGCGC CGCAATACGA
50 951 CATCACCCCC ATCGATCCGT CCGGCCTGAA GCAGCAGTCG GCAGGCAATC
1001 TGTTTAAATT GGAATACGAC GCGGTATTC AATAATACAC GGCGCAATTT
1051 CGCGATTTAA ACACAGAAAT CGGCAGCCG AAAATCATCA ACCGCAATTA
1101 TCAATTCAT TACGTTTGT CTTTGAACCC GTATACCAAC CTCATCTGA
1151 CCGCAGCCTA CAATTCGGGC AGGCAGAAAT ATCCGAAAGG GGCGAAGTTT
55 1201 ACAGGCTGGG GGCTTTTAAA AGATTTTGAA ACCTACAACA ACGCGAAAAT
1251 CCTCGACCTC AACAACACCG CCACCTCCG GCTGCCCGC GAAACCGAGT
1301 TGCAAACAC TTTGGGCTTC AATTATTTC ACAACGAATA CGGCAAAAC
1351 CGCTTTCCTG AAGAATTGGG GCTGTTTTC GACGGTCCTG ATCAGGACAA
1401 CGGGCTTTAT TCCTATTGG GCGGTTTAA GGGCGATAAA GGGCTGTTGC
60 1451 CTCAAAAATC AACCATTGTC CAACCGGCGG GCAGCAATA TTTCACACG
1501 TTCTACTTCG ATGCCGCGCT CAAAAAGAC ATTTACCGCT TAACTACAG
1551 CACCAATGCA ATCAACTACC GTTTCGGCGG CGAATATACG GGCTATTACG
1601 GCTCGGAAAA CGAATTTAAG CGGGCATTTC GAGAAAACTC GCCGGCATAC
1651 AAGGAACATT GCGACCCGAG CTGCGGGCTT TATGAACCCG TATTGAAAAA
65 1701 ATACGGCAAA AAGCGCGCCA ACAACCATC GGTCAGCATT AGTGCGGACT
1751 TCGCGGATTA TTTCATGCCG TTCGCGGGCT ATTCGCGCAC ACACCGTATG

```



5	1851	CACCGCCTTA	AAACCAGAGC	GCGCAAACAC	TTGGCAATTT	GGCTCAATA
	1901	CCTATAAAAA	AGGATTGTTA	AAACAAGATG	ATATATTAGG	ATTGAAACTG
	1951	GTCCGGCTACC	GCAGCCGCAT	TGACAACTAC	ATCCACAACG	TTTACGGGAA
	2001	ATGGTGGGAT	TTGAACGGGG	ATATTCCGAG	CTGGGTCCGC	AGCACCGGGC
	2051	TTGCCTACAC	CATCCGACAC	CGCAATTTCA	AAGACAAAGT	GCACAAACAC
10	2101	GGTTTTGAGC	TGGAGCTGAA	TTACGATTAT	GGGCGTTTTT	TCACCAACCT
	2151	TTCTTACGCC	TATCAAAAAA	GCACGCAACC	GACCAATTTT	AGCGATGCGA
	2201	GGCAATCGCC	CAACAATGCC	tccaaAGAAG	ACCAACTCAA	ACAAGGTTAT
	2251	CGGCTGAGCA	GGGTTTCCGC	CCTGCCCGCA	GATTACGGAC	GTTTGGAAGT
	2301	CGGTACGCGC	TGGTTGGGCA	ACAAACTGAC	TTTGGGCGGC	GCGAtgcGCT
15	2351	ATTTCCGGCAA	GAGCATCCGC	GCGACGGCTG	AAGAACGCTA	TATCGACGGC
	2401	ACCAACGGGG	GAAATCCACG	CAATGTCCGG	CAACTGGGCA	AGCGTTCCAT
	2451	CAAAACAAAC	GAAACCCTTG	CCCACAGCC	TTTGATTTTT	GATTTTTACG
	2501	CCGCTTACGA	GCCGAAGAAA	AACCTTATTT	TCCGCGCCGA	AGTCAAAAAC
	2551	CTGTTCGACA	GGCGTTATAT	CGATCCGCTC	GATGCGGGCA	ATGATGCGGC
20	2601	AACGCAGCGT	TATTACAGCT	CGTTCGACCC	GAAAGACAAG	GACGAAGACG
	2651	TAACGTGTAA	TGCTGATAAA	ACGTTGTGCA	ACGGCAAAAT	CGGCGGCACA
	2701	AGCAAAAGCG	TATTGACCAA	TTTCGCACGC	GGACGCACCT	TCTTGATGAC
	2751	GATGAGCTAC	AAGTTTTTAA			

This corresponds to the amino acid sequence <SEQ ID 884; ORF133ng-1>:

	1	MRSSFRLKPI	CFYLMGVMLY	HHSYAEDAGR	AGSEAQIQVL	EDVHVKAKRVR
	51	PKDKKVFTDA	RAVSTRQDVF	KSGENLDNIV	RSIPGAFTQQ	DKSSSGIVSLN
25	101	IRGDSGFGRV	NTMVDGITQT	FYSTSTDAGR	AGGSSQFGAS	VDSNFIAGLD
	151	VVKGSFSGSA	GINSLAGSAN	LRTLGVDDVV	QGNNTYGLLL	KGLTGTNSTK
	201	GNAMAAIGAR	KWLESGASVG	VLYGHSRRGV	AQNYRVGGG	QHIGNFGEEY
	251	LERRKQQYFV	QEGGLKFNAG	SGKWERDLQR	QYWKTKWYKK	YEDPQELQKY
	301	IEEHDKSWRE	NLAPQYDITP	IDPSGLKQOS	AGNLFKLEYD	GVFNKYTAQF
30	351	RDLNTRIGSR	KIINRNYQFN	YGLSLNPYTN	LNLTAAYNSG	RQKYPKGAKF
	401	TGWLLKDFE	TYNNAKILDL	NNTATFRLPR	ETELQTTLGF	NYFHNIEYGN
	451	RFPEELGLFF	DGPDQDNGLY	SYLGRFKGDK	GLLPQKSTIV	QPAGSQYFNT
	501	FYFDAALKKD	IYRLNYSTNA	INYRFGGEYT	GGYGSSENEF	RAFGENSPAY
	551	KEHCDPSCGL	YEPVLKKYK	KRANNHSVSI	SADFGDYFMP	FAGYSRTHRM
35	601	PNIQEMYFSQ	IGDSGVHTAL	KPERANTWQF	GFNTYKKGLL	KQDDILGLKL
	651	VGYRSRIDNY	IHNVYGKWWD	LNGDIPSWVG	STGLAYTIRH	RNFKDKVHKH
	701	GFELELNIDY	GRFFTNLSYA	YQKSTQPTNF	SDASESPNNA	SKEDQLKQGY
	751	GLSRVSALPR	DYGRLEVQTR	WLGNKLTGG	AMRYFGKSIR	ATAEERYIDG
	801	TNGGRTSNVR	QLGKRSIKQT	ETLARQPLIF	DFYAAYEPKK	NLIFRAEVKN
40	851	LFDRRYIDPL	DAGNDAATOR	YSSSFDPKDK	DEDVTCNADK	TLCNGKYGGT
	901	SKSVLTNEAR	GRTFLMTMSY	KF*		

ORF133ng-1 and ORF133-1 show 96.2% identity in 889 aa overlap:

		10	20	30	40	50	60
45	orf133ng-1.pep	SFRLKPICFYLMGVMLYHHSYAEDAGRAGSE	EAQIQVLEDVHV	KAKRV	PKDKKV	FTDARAV	
	orf133-1				EAQIQVLEDVHV	KAKRV	PKDKKV
					10	20	30
50	orf133ng-1.pep	STRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGRVNTMVDGITQTFYS					
	orf133-1	STRQDIFKSSSENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGRVNTMVDGITQTFYS					
		40	50	60	70	80	90
55	orf133ng-1.pep	TSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVQGN					
	orf133-1	TSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVQGN					
		100	110	120	130	140	150
60	orf133ng-1.pep	NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLVYGH					
	orf133-1	NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLVYGH					
		160	170	180	190	200	210
65	orf133ng-1.pep	GNFGEEYLERRKQQYFVQEGGLKFNAGSGKWERDLQRQYWKTKWYKKYEDPQELQKYIEE					
		250	260	270	280	290	300

-481-

```

||||| |||||:|||||: ||||| ||| |||:|: |||||
orf133-1  GNFGAEYLERRKQRYFVQEGALKFNSSDQKWERDLQRQWKYKPYKNNN-QELQKYEIE
          220      230      240      250      260

5
          310      320      330      340      350      360
orf133ng-1.pep HDKSWRENLAPOYDITPIDPSGLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII
          ||||| |||||:|||||: ||||| |||||: |||||
orf133-1      HDKSWRENLAPOYDITPIDPSGLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII
          270      280      290      300      310      320

10
          370      380      390      400      410      420
orf133ng-1.pep NRNYQFNGLSLNPYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDNLNT
          ||||| |||||:|||||: ||||| |||||: |||||
orf133-1      NRNYQFNGLSLNPYTNLNLTAAYNSGRQKYPKGSKFTGWGLLKDFETYNNAKILDNLNT
          330      340      350      360      370      380

15
          430      440      450      460      470      480
orf133ng-1.pep ATFRLPRETELQTTLGFNFYHNEYGKNRFEELGLFFDGPQDNGLYSYLGRFKGDKGLL
          ||||| |||||:|||||: ||||| |||||: |||||
orf133-1      ATFRLPRETELQTTLGFNFYHNEYGKNRFEELGLFFDGPQDNGLYSYLGRFKGDKGLL
          390      400      410      420      430      440

20
          490      500      510      520      530      540
orf133ng-1.pep PQKSTIVQAGSQYFNTFYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKRAF
          ||||| |||||:|||||: ||||| |||||: |||||
orf133-1      PQKSTIVQAGSQYFNTFYFDAALKKDIYRLNYSTNTVGYRFGGEYTGYYGSDDEFKRAF
          450      460      470      480      490      500

25
          550      560      570      580      590      600
orf133ng-1.pep GENSPAYKEHCDPSCGLYEPVLKKYGKKRANNHVSISADFGDYFMPFAGYSRTHRMPNI
          ||||| |||||:|||||: ||||| |||||: |||||
orf133-1      GENSPYKHKCNRSCGIYEPVLKKYGKKRANNHVSISADFGDYFMPFASYSRTHRMPNI
          510      520      530      540      550      560

30
          610      620      630      640      650      660
orf133ng-1.pep QEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDILGLKLGVYRSRIDNYIHN
          ||||| |||||:|||||: ||||| |||||: |||||
orf133-1      QEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDTLGLKLGVYRSRIDNYIHN
          570      580      590      600      610      620

35
          670      680      690      700      710      720
orf133ng-1.pep VYGKWWDLNGDIPSWVGSTGLAYTIHRNFKDKVHKHGFELNLDYGRFFTNLSYAYQK
          ||||| |||||:|||||: ||||| |||||: |||||
orf133-1      VYGKWWDLNGDIPSWVSSTGLAYTIQHRNFKDKVHKHGFELNLDYGRFFTNLSYAYQK
          630      640      650      660      670      680

40
          730      740      750      760      770      780
orf133ng-1.pep STQPTNFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLNKLTGGAMR
          ||||| |||||:|||||: ||||| |||||: |||||
orf133-1      STQPTNFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLNKLTGGAMR
          690      700      710      720      730      740

45
          790      800      810      820      830      840
orf133ng-1.pep YFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDFYAAAYEPKKNLI
          ||||| |||||:|||||: ||||| |||||: |||||
orf133-1      YFGKSIRATAEERYIDGTNGGNTSNFRQLGKRSIKQTETLARQPLIFDFYAAAYEPKKNLI
          750      760      770      780      790      800

50
          850      860      870      880      890      900
orf133ng-1.pep FRAEVKNLFDRLYIDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLNKGKYGTSKS
          ||||| |||||:|||||: ||||| |||||: |||||
orf133-1      FRAEVKNLFDRLYIDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLNKGKYGTSKS
          810      820      830      840      850      860

55
          910      920
orf133ng-1.pep VLTNFARGRTFLMTMSYKFX
          |||||
orf133-1      VLTNFARGRTFLMTMSYKFX
          870      880

```

70 In addition, ORF133ng-1 is homologous to a TonB-dependent receptor in *H.influenzae*:

-482-

sp|P45114|YC17\_HAEIN PROBABLE TONB-DEPENDENT RECEPTOR HI1217 PRECURSOR  
>gi|1075372|pir|G64110 transferrin binding protein 1 precursor (tbpl) homolog -  
Haemophilus influenzae (strain Rd KW20) >gi|1574147 (U32801) transferrin binding  
protein 1 precursor (tbpl) [Haemophilus influenzae] Length = 913  
Score = 930 bits (2377), Expect = 0.0  
Identities = 476/921 (51%), Positives = 619/921 (66%), Gaps = 72/921 (7%)

Query: 38 QVLEDVHVKAARVPKDKKVFTDARAVSTRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIV 97  
+ L + V K + DKK FT+A+A STR++VFK + +D ++RSIPGAFTQQDK SG+V  
Sbjct: 29 ETLGQIDVVEKVISNDKKPFTEAKAKSTRENVFKETQTIDQVIRSIPGAFTQQDKSGSVV 88

Query: 98 SLNIRGDSGFRVNTMVDGITQTFYSTSTADAGRAGSSQFGASVDSNFIAGLDVVKGSFS 157  
S+NIRG++G GRVNTMVDG+TQTFYST+ D+G++GGSSQFGA++D NFIAG+DV K +FS  
Sbjct: 89 SVNIRGENGLGRVNTMVDGVTQTFYSTALDSGQSGSSQFGAIDPNFIAGVDVNKSNSFS 148

Query: 158 GSAGINSLAGSANLRTLGVDDVVQXXXXXXXXXXXXXXXXXXXXXAMAAIGARKWLESGA 217  
G++GIN+LAGSAN RTLGV+DV+ M RKWL++G  
Sbjct: 149 GASGINALAGSANFRTLGVNDVITDDKPFGIILKGMTGSNATKSNFTMAAGRKWLDNGG 208

Query: 218 SVGVLYGHSRRGVAQNYRVGGGQHHIGNEEYLERRKQYFVQEGGLKFNAGSGKWERD 277  
VGV+YG+S+R V+Q+YR+ GGG+ + + G++ L + K+ YF + G N G+W D  
Sbjct: 209 YGVVYGYQSREVSQDYRI-GGGERLASLGQDILAKEKEAYF-RNAGYILNP-EGQWTPD 265

Query: 278 LQRQYWK-----TKWY-----KQYEDPQELQK---YIEE 303  
L +++W +Y KK +D ++LQK IEE  
Sbjct: 266 LSKKHWSCKNKPDPYQKNGDCSYRIGSAAKTRREILQELLTNGKKPKDIEKLQKNGDGIEE 325

Query: 304 HDKSWRENLAPOYDITPIDPSGLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII 363  
DKS+ N QY + PI+P L+ +S +L K EY AQ R L+ +IGSRKI  
Sbjct: 326 TDKSFERN-KDQYSVAPIEPGSLQSRSRSHLLKFEYGDHQNGLAQLRTLNDKIGSRKIE 384

Query: 364 NRNYQFNYSLSLNPYTNLNLTAAYNSGRQKYPKGAFTGWGLLKDFETYNNAKILDNLNT 423  
NRNYQ NY + N Y +LNL AA+N G+ YPKG F GW + T N A I+D+NN+  
Sbjct: 385 NRNYQVNYNFNNNSYLDLNLMAAHNIGKTIYPKGGFFAGWQVADKLITKNVANIVDINNS 444

Query: 424 ATFRLPRETELQTTLGFNYPHNEYGKNRFEELGLFFDGPQDQNGLYSY--LGRFKGDKG 481  
TF LP+E +L+TTLGFNYP NEY KNRFEEL LF++ D GLYS+ GR+ G K  
Sbjct: 445 HTFLLPKEIDLKTTLGFNYPFTNEYSKNRFEELSIFYNDASHDQGLYSHSKRGRYSGTKS 504

Query: 482 LLPQKSTIVQFAGSQYFNTFYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKR 541  
LLPQ+S I+QP+G Q F T YFD AL K IY LNYS N +Y F GEY GY  
Sbjct: 505 LLPQRSVILQPSGKQKFKTVYFDTALSKGIYHLNYSVNFTHYAFNGEYVGY----- 555

Query: 542 AFGENSPAYKEHCDPSCGLYEPVLKKYKGRANNHVSISADFGDYFMPFAGYSRTHRMP 601  
EN+ + EP+L K G K+A NHS ++SA+ DYEMPF YSRTHRMP  
Sbjct: 556 ---ENTAGQQ-----INEPILHKS GHKKA FNHSATLSAELSDYFMPFFTYSRTHRMP 604

Query: 602 NIQEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYI 661  
NIQEM+FSQ+ ++GV+TALKPE+++T+Q GFNTYKKGL QDD+LG+KLVGYRS I NYI  
Sbjct: 605 NIQEMFFSQVSNAGVNTALKPEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFIKNYI 664

Query: 662 HNVYKWWDLNGDIPSWVGSTGLAYTIRHRNFKDKVHKHGFLELNYDYGRFFTNLSYAY 721  
HNVYG WW +P+W S G YTI H+N+K V K G ELE+NYD GRFF N+SYAY  
Sbjct: 665 HNVYGVVW--RDGMPTWAESNGFKYITIAHQNYKPIVKKSGVELEINYDMGRFFANVSYAY 722

Query: 722 QKSTQPTNFSDESPPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGLGA 781  
Q++ QPTN++DAS PNNAS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTGL A  
Sbjct: 723 QRTNQPTNYADASPRPNNASQEDILKQGYGLSRVSMLPKDYGRLELGTWFDQKLTGLA 782

Query: 782 MRYFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDYAAEPKKN 841  
RY+GKS RAT EE YI+G+ + +R+ ++K+TE + +QP+I D + +YEP K+  
Sbjct: 783 ARYYGKSKRATIEEYINGSR-FKKNTLRRENYAVKKTEDIKKQPIILDHVSYPEIKD 841

Query: 842 LIFRAEVKNLFDRRYIDPLDAGNDAATQRYSSSFDPKDKDEDTVCNADKTLGNGKYGGTS 901  
LI +AEV+NL D+RY+DPLDAGNDA+QRYSS + + C D + C GG+  
Sbjct: 842 LIIKAEVQNLLDKRYVDPLDAGNDAASQRYSSS-----NNSIECAQDSSAC----GGSD 892

Query: 902 KSVLTNFARGRTFLMTMSYKF 922  
K+VL NFARGRT+++++YKF  
Sbjct: 893 KTVLYNFARGRTYILSLNYKF 913

The underlined motif in the gonococcal protein (also present in the meningococcal protein) is predicted to be an ATP/GTP-binding site motif A (P-loop), and the analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 5 Example 104

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 885>

```

1 ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
51 TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG
10 151 GGCTACACCG CCCTCAAAAT GCCCGCCCGC GCCTACGAAC TGATTCCCCT
201 CGCCGTCCTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCCGGCA
251 GCGAACTGAC CGTCATCAAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
301 TTGATTCTGT CGCAGTTCGG TTTTATTTTT GCTATTGCCA CCGTCGCGCT
351 CGGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG
15 401 CCGCCGCCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
451 AAAGAAAAAA ACAGCGTGAT CAATGTGCGC GAAATGTTGC CCGACCAT..

```

This corresponds to the amino acid sequence <SEQ ID 886; ORF112>:

```

1 MNLISRYIIR QMAVMAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML
51 GYTALKMPAR AYELIPLAVL IGGLVLSLSQL AAGSELTVIK ASGMSTKKLL
20 101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL
151 KEKNSVINVR EMLPDH...

```

Further work revealed further partial nucleotide sequence <SEQ ID 887>:

```

1 ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
51 TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
25 101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG
151 gGCTACACCG CCCTCAAAAT GCCCGCCCGC GCCTACGAAC TGATTCCCCT
201 CGCCGTCCTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCCGGCA
251 GCGAACTGAC CGTCATCAAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
301 TTGATTCTGT CGCAGTTCGG TTTTATTTTT GCTATTGCCA CCGTCGCGCT
30 351 CGGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG
401 CCGCCGCCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
451 AAAGAAAAAA ACAGCTKAT CAATGTGCGC GAAATGTTGC CCGACCATAC
501 GCTTTTGGGC ATCAAAATTT GGGCGCGCAA CGATAAAAC GAATTGGCAG
551 AGGCAGTGGA AGCCGATTCC GCCGTTTGA ACAGCGACGG CAGTTGGCAG
35 601 TTGAAAAACA TCCGCCGCAG CACGCTTGGC GAAGACAAAG TCGAGGTCTC
651 TATTGCGGCT GAAGAAACT GGCCGATTTC CGTCAAACGC AACCTGATGG
701 ACGTATTGCT CGTCAAACCC GACCAAATGT CCGTCGGCGA ACTGACCACC
751 TACATCCGCC ACCTCCAAAA CAACAGCCAA AACACCCGAA TCTACGCCAT
801 CGCATGGTGG CGCAAATTGG TTTACCCCGC CGCAGCCTGG GTGATGGCGC
40 851 TCGTCGCCTT TGCCTTACC CCGCAAACCA CCCGCCACGG CAATATGGGC
901 TTAATACTCT TCGCGGCAT CTGTsTCGGA TTGCTGTTCC ACCTTGCCGG
951 ACGGCTCTTT GGGTTACCA GCCAACTCGG...

```

This corresponds to the amino acid sequence <SEQ ID 888; ORF112-1>:

```

1 MNLISRYIIR QMAVMAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML
45 51 GYTALKMPAR AYELIPLAVL IGGLVLSLSQL AAGSELTVIK ASGMSTKKLL
101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL
151 KEKNSXINVR EMLPDHTLLG IKIWARNDKN ELAEAVEADS AVLNSDGSWQ
201 LKNIRRTLGL EDKVEVSIAA EENWPISVKR NLMDVLLVKP DQMSVGLTST
251 YIRHLQNNNSQ NTRIYAIWW RKLVPAAAW VMALVAFAFT POTTRHGNMG
50 301 LKLFGGICXG LLFHLAGRLF GFTSQL...

```

Computer analysis of this amino acid sequence predicts two transmembrane domains and gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF112 shows 96.4% identity over a 166aa overlap with an ORF (ORF112a) from strain A of *N.*

*meningitidis*:

5	orf112.pep	10	20	30	40	50	60
	orf112a	10	20	30	40	50	60
10	orf112.pep	70	80	90	100	110	120
	orf112a	70	80	90	100	110	120
15	orf112.pep	130	140	150	160		
	orf112a	130	140	150	160	170	180
20	orf112a	130	140	150	160	170	180
	orf112a	190	200	210	220	230	240

The ORF112a nucleotide sequence <SEQ ID 889> is:

25	1	ATGAACCTGA	TTTCACGTTA	CATCATCCGT	CAATGGCGG	TTATGGCGGT
	51	TTACGCGCTC	CTTGCCCTTC	TCGCTTTGTA	CAGCTTTTTT	GAAATCCTGT
30	101	ACGAAACCGG	CAACCTCGGC	AAAGGCAGTT	ACGCATATG	GGAAATGNTG
	151	GGNTACACCG	CCCTCAAAT	GNCCGCCGC	GCCTACGAAC	TGATGCCCCT
35	201	CGCCGTCCTT	ATCGGCGGAC	TGGTCTCTNT	CAGCCAGCTT	GCCGCCGGCA
	251	GCGAACTGAN	CGTCATCAA	GCCAGCGGCA	TGAGACCAA	AAAGCTGCTG
40	301	TTGATTCTGT	CGCAGTTCGG	TTTTATTTT	GCTATTGCCA	CCGTGCGGCT
	351	CGGCGAATGG	GTTGCGCCA	CACTGAGCCA	AAAAGCCGAA	AACATCAAAG
45	401	CCGCGGCCAT	CAACGGCAAA	ATCAGTACCG	GCAATACCGG	CCTTTGGCTG
	451	AAAGAAAAA	ACAGCATTAT	CAATGTGCGC	GAAATGTTGC	CCGACCATAC
50	501	CCTGCTGGGC	ATTAAATCT	GGGCCCGCAA	CGATAAAAC	GAAGTGGCAG
	551	AGGCAGTGA	AGCCGATTCC	GCCGTTTGA	ACAGCGACGG	CAGTTGGCAG
55	601	TTGAAAAACA	TCCGCCGAG	CACGCTGGC	GAAGACAAAG	TCGAGGTCTC
	651	TATTGCGGCT	GAAGAAANT	GGCCGATTTC	CGTCAAACGC	AACCTGATGG
60	701	ACGTATTGCT	CGTCAAACCC	GACCAAATGT	CCGTGCGCGA	ACTGACCACC
	751	TACATCCGCC	ACCTCCAAAN	NNACAGCCAA	AACACCCGAA	TCTACGCCAT
65	801	CGCATGGTGG	CGCAAATTGG	TTTACCCCGC	CGCAGCCTGG	GTGATGGCGC
	851	TCGTGCGCTT	TGCCTTTACC	CCGCAAACCA	CCCGCCACGG	CAATATGGGC
70	901	TTAAANTCT	TCGGCGGCAT	CTGTCTCGGA	TTGCTGTTCC	ACCTTGCCGG
	951	NCGGCTCTTC	NGGTTTACCA	GCCAACTCTA	CGGCATCCCG	CCCTTCCTCG
75	1001	NCGGCGCACT	ACCTACCATA	GCCTTCGCCT	TGCTCGCCGT	TTGGCTGATA
	1051	CGCAAACAGG	AAAAACGCTA	A		

This encodes a protein having the amino acid sequence <SEQ ID 890>:

50	1	MNLISRYIIR	QMAVMAYAL	LAFLALYSFF	EILYETGNLG	KGSYGIWEMX
	51	GYTALKMXAR	AYELMPLAVL	IGGLVSXSQ	AAGSELXVIK	ASGMSTKKLL
55	101	LILSQFGFIF	AIATVALGEW	VAPTLSQKAE	NIKAAINGK	ISTGNTGLWL
	151	KEKNSIINVR	EMLPDHTLLG	IKIWARNDKN	ELAEAVEADS	AVLNSDGSWQ
60	201	LKNIRRLSTL	EDKVEVSIAA	EEXWPISVVR	NLMDVLLVKP	DQMSVGLT
	251	YIRHLQXXSQ	NTRIYAIWW	RKLVPAAAW	VMALVAFAFT	PQTRHGNMG
65	301	LKXFGGICLG	LLFHLAARLF	XFTSOLYGIP	PFLXGALPTI	AFALLAVWLI
	351	RKQEK*				

ORF112a and ORF112-1 show 96.3% identity in 326 aa overlap:

60	orf112a.pep	MNLISRYIIRQMAVMAYAL	LAFLALYSFF	EILYETGNLG	KGSYGIWEMX	GYTALKMXAR
	orf112-1	MNLISRYIIRQMAVMAYAL	LAFLALYSFF	EILYETGNLG	KGSYGIWEMX	GYTALKMXAR
	orf112a.pep	AYELMPLAVLIGGLVSXSQ	AAGSELXVIK	ASGMSTKKLL	LILSQFGFIF	AIATVALGEW

	orf112-1	AYELIPLAVLIGGLVSLSQLAAGSELTVIKASGMSTKKLLILSQFGFIFAIATVALGEW
5	orf112a.pep	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSIINVREMLPDHTLLGIKIWARNDKN
	orf112-1	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSXINVREMLPDHTLLGIKIWARNDKN
	orf112a.pep	ELAEAVEADSAVLNSDGSWQLKNIRRSTLGEDKVEVSIAAEEXWPISVKRNLMDEVLLVKP
10	orf112-1	ELAEAVEADSAVLNSDGSWQLKNIRRSTLGEDKVEVSIAAEENWPISVKRNLMDEVLLVKP
	orf112a.pep	DQMSVGELTTYIRHLQXXSQNTRIYAIAWWRKLVYPAAAWVMALVAFAFTPQTTRHGNMG
	orf112-1	DQMSVGELTTYIRHLQNNSQNTRIYAIAWWRKLVYPAAAWVMALVAFAFTPQTTRHGNMG
15	orf112a.pep	LKXFGGICLGLLFHLAGRLFXFTSOLYGIPPFLXGALPTIAFALLAVWLIRKQEKRX
	orf112-1	LKLFGGICXGLLFHLAGRLFGFTSQL

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF112 shows 95.8% identity over 166aa overlap with a predicted ORF (ORF112ng) from *N.*

*gonorrhoeae:*

25	orf112.pep	MNLISRYIIRQMAVMNAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR	60
	orf112.ng	MNLISRYIIRQMAVMNAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR	60
30	orf112.pep	AYELIPLAVLIGGLVLSQLAAGSELTVIKASGMSTKKLLLLILSQFGFIFAIATVALGEW	120
		:     :     :     :     :     :     :	
	orf112.ng	AYELMPLAVLIGGLASLSQLAAGSELAVIKASGMSTKKLLLLILSQFGFIFAIAAVALGEW	120
	orf112.pep	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSVINVREMLPDH	166
		:     :     :     :     :     :	
	orf112.ng	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKTSIINVRGMLPDHTLLGIKIWARNDKN	180

The complete length ORF112ng nucleotide sequence <SEQ ID 891> is:

35	1	ATGAACCTGA	TTTCACGTTA	CATCATCCGC	CAAATGGCGG	TTATGGCGGT
	51	TTACGCGCTC	CTTGCCCTTC	TCGCTTTGTA	CAGCTTTTTT	GAATCCTGT
	101	ACGAAACCGG	CAACCTCGGC	AAAGGCAGTT	ACGGCATATG	GGAAATGCTG
	151	GGCTACACCG	CCCTCAAAT	GCCGCGCCGC	GCCTACGAAC	TCATGCCCTT
40	201	CGCCGTCTCT	ATCGGCGGAC	TGGCTCTCT	CAGCCAGCTT	GCCGCGGCA
	251	CGCAACTGGC	CGTCATCAAA	GCCAGCGGCA	TGAGCACCAA	AAGAGCTGCTG
	301	TTGATTCTGT	CTCAGTTCGG	TTTTATTTTT	GCTATTGCCG	CCGTGCGGCT
	351	CGGCGAATGG	GTTGCGCCCA	CGCTGAGCCA	AAAAGCCGAA	AACATCAAag
45	401	cCGCCGCCAt	taacgGCAAA	ATCAGCAcgg	gcAATACCGG	CCTTTggcTG
	451	AAAGAAAAAA	ccAGCATTAT	CAATGTGcGc	GGAATGTTGC	CCGACCATAC
	501	GCTTTTGGGC	ATCAAAATTT	GGGCGCGCAA	CATATAAAAC	GAATTGGCAG
	551	AGGCAGTGGa	AGCCGATTCC	GCCGTTTTGA	ACAGCGACGG	CAGCTGGCAG
50	601	TTGAAAAACA	TCCGCCGCAG	CATCATGGGT	ACAGACAAAA	TCGAAACATC
	651	cgCCGCCGCC	GAAGAAACTT	gGCCGATTGC	CGTCAGACGC	AACCTGATGG
	701	ACGTATTGCT	CGTCAAGCCC	GACCAATGT	CCGTCGGCGA	GCTGACCACC
	751	TACATCCGGC	ACCTCAAAA	CAACAGCCAA	AACACCCAAA	TCTACGCCAT
55	801	CGCATGGTGG	CGTAAACTCG	TTTACCCCGT	CGCCGCATGG	GTCAATGGCGC
	851	TCGTTGCCTT	CGCCTTTACG	CCGCAAACCA	CGCGCCACGG	CAATATGGGC
	901	TTAAAACCTCT	TCGGCGGCAT	CTGTCTCGGA	TTGCTGTTCC	ACCTTGCCCG
	951	CAGGCTCTTC	GGGTTTACCA	GCCAACTCTA	CGGCACCCCA	CCCTTCCTCG
	1001	CCGGCGCACT	GCCTACCATa	GCCTTCGCCT	TGCTCGCTGT	TTGGCTGATA
	1051	CGCAAAACAGG	AAAAACGTTG	A		

This encodes a protein having amino acid sequence <SEQ ID 892>:

	1	<u>MNLISRYIIR</u>	<u>QMAVMVAYAL</u>	<u>LAFLALYSFF</u>	<u>EILYETGNLG</u>	<u>KGSYGIWEML</u>
	51	<u>GYTALKMPAR</u>	<u>AYELMPLAVL</u>	<u>IGGLASLSQL</u>	<u>AAGSELAVIK</u>	<u>ASGMSTKKLL</u>
60	101	<u>LILSQFGFIF</u>	<u>AIAAVALGEW</u>	<u>VAPTLSQKAE</u>	<u>NIKAAAINGK</u>	<u>ISTGNTGLWL</u>
	151	<u>KEKTSIINVR</u>	<u>GMLPDHTLLG</u>	<u>IKIWARNDKN</u>	<u>ELAEAVEADS</u>	<u>AVLNSDGSWQ</u>
	201	<u>LKNIRRSIMG</u>	<u>TDKIETSAAS</u>	<u>EETWPIAVRR</u>	<u>NLMDVLLVKP</u>	<u>DQMSVGEILT</u>
	251	<u>YIRHLQNNSQ</u>	<u>NQIYAIAWW</u>	<u>RKLVPVPAAR</u>	<u>VMALVEAFET</u>	<u>PQTRRHGNMG</u>
	301	<u>LKLEGGICLG</u>	<u>LLFHLAGRLF</u>	<u>GFTSOLYGTP</u>	<u>PFLAGALPTI</u>	<u>AFALLAVWLI</u>

351 RKQEK\*  
RKQEK\*

ORF112ng and ORF112-1 show 94.2% identity in 326 aa overlap:

		10	20	30	40	50	60
5	orf112ng	MNLISRYII	QMAVM	AVYALLA	FLALYS	SFFEILY	ETGNLGKGSYGIWEM
	orf112-1	MNLISRYII	QMAVM	AVYALLA	FLALYS	SFFEILY	ETGNLGKGSYGIWEM
		10	20	30	40	50	60
10	orf112ng	70	80	90	100	110	120
		AYELMPLA	VLIGGLA	SLSQLAA	GSELAVI	KASGMST	KKLLILS
	orf112-1	AYELIPLA	VLIGGLV	SLSQLAA	GSELTVI	KASGMST	KKLLILS
		70	80	90	100	110	120
15	orf112ng	130	140	150	160	170	180
		VAPTLSQ	KAENIKA	AAAINGK	ISTGNTG	LWLKEKT	SIINVRG
	orf112-1	VAPTLSQ	KAENIKA	AAAINGK	ISTGNTG	LWLKEKNS	XINVREML
		130	140	150	160	170	180
20	orf112ng	190	200	210	220	230	240
		ELAEAVEA	DSAVLNS	DGSGWL	KNIRRSI	MGTDKI	ETSAAAE
	orf112-1	ELAEAVEA	DSAVLNS	DGSGWL	KNIRRST	LGEDKVE	VSIAAEN
		190	200	210	220	230	240
25	orf112ng	250	260	270	280	290	300
		DQMSVGEL	TTYIRHL	QNNNQNT	QIYAI	AWWRKL	VYPVAAW
	orf112-1	DQMSVGEL	TTYIRHL	QNNNQNT	RIYAI	AWWRKL	VYPVAAW
		250	260	270	280	290	300
30	orf112ng	310	320	330	340	350	
		LKLFGGIC	LGLLFL	AGRLFG	FTSQLY	GTTPFLA	GALPTIA
	orf112-1	LKLFGGIC	XGLLFL	AGRLFG	FTSQL		
		310	320				

This analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their  
 40 epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

It will be appreciated that the invention has been described by means of example only, and that  
 modifications may be made whilst remaining within the spirit and scope of the invention.

TABLE I – PCR primers

ORF	Primer	Sequence	Restriction sites
ORF 1	Forward	CGCGGATCCGCTAGC-GGACACACTTATTTTCGG	BamHI-NheI
	Reverse	CCCGCTCGAG-CCAGCGGTAGCCTAATT	XhoI
ORF 2	Forward	GCGGATCCCATATG-TTTGATTTCGGTTTGGG	BamHI-NdeI
	Reverse	CCCGCTCGAG-GACGGCATAACGGCG	XhoI
ORF 2-1	Forward	GCGGATCCCATATG-TTTGATTTCGGTTTGGG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TGATTTACGGACGCGCA	XhoI
ORF 4	Forward	GCGGATCCCATATG-TGCGGAGGTCAAAAAGAC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTGGCTGCGCCTTC	XhoI
ORF 5	Forward	GGAATTCATATGGCCATGG-TGGAAGGCGCACAACC	NdeI-NcoI
	Forward	CGGGATCC-ATGGAAGGCGCACAAC	BamHI
	Reverse	CCCGCTCGAG-GACTGTGCAAAAACGG	XhoI
ORF 6	Forward	CGCGGATCCCATATG-ACCCGTCAATCTCTGCA	BamHI-NdeI
	Reverse	CCCGCTCGAG-TGCGCCGAACACTTTC	XhoI
ORF 7	Forward	CGCGGATCCGCTAGC-GCGCTGCTTTTGTTC	BamHI-NheI
	Reverse	CCCGCTCGAG-TTTCAAAATATATTTGCGGA	XhoI
ORF 8	Forward	GCGGATCCCATATG-GCTCAACTGCTTCGTAC	BamHI-NdeI
	Reverse	CCCGCTCGAG-AGCAGGCTTTGGCGC	XhoI
ORF 9	Forward	CGCGGATCCCATATG-CCGAAGGAAGTCGGAAA	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTCGAGGTTTTCGGG	XhoI
ORF 10	Forward	GCGGATCCCATATG-GACACAAAAGAAATCCTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAATGGGAAACCTTGTTTT	XhoI
ORF 11	Forward	GCGGATCCCATATG-GCGGTCAACCTCTACG	BamHI-NdeI
	Reverse	CCCGCTCGAG-GGAAACGACTTCGCC	XhoI
ORF 13	Forward	CGCGGATCCCATATG-GCTCTGCTTTCCGCGC	BamHI-NdeI
	Reverse	CCCGCTCGAG-AGGGTGTGTGATAATAAG	XhoI
ORF 15	Forward	GGAATTCATATGGCCATGG-GCGGGACACTGACAG	NdeI-NcoI
	Forward	CGGGATCC-TGCGGGACACTGACAGG	BamHI
	Reverse	CCCGCTCGAG-AGGTTGGCCTTGCTATG	XhoI
ORF 17	Forward	GGAATTCATATGGCCATGG-TTGCCGGCCTGTTCG	NdeI-NcoI



	Forward Reverse	<u>CGGGATCC</u> -ATTGCCGGCCTGTTTCG CCCGCTCGAG-AAGCAGGTTGTACAGC	BamHI XhoI
ORF 18	Forward Reverse	<u>GCGGATCCCATATG</u> -ATTTTGCTGCATTTGGAT CCCGCTCGAG-TCTTCAATTTCTGAAAGC	BamHI-NdeI XhoI
ORF 19	Forward Forward Reverse	GGAATTC <u>CATATGGCCATGG</u> -TCGCCAGTGTTTTTACC <u>CGGGATCC</u> -TTCGCCAGTGTTTTTACCG CCCGCTCGAG-GGTGTTTTTGAAGCTGCC	NdeI-NcoI BamHI XhoI
ORF 20	Forward Forward Reverse	GGAATTC <u>CATATGGCCATGG</u> -TCGGCGCGGGTATG <u>CGGGATCC</u> -TTCGGCGCGGGTATG CCCGCTCGAG-CGGCGAGCGAGAGCA	NdeI-NcoI BamHI XhoI
ORF 22	Forward Forward Reverse	GGAATTC <u>CATATGGCCATGG</u> -TGATTAAATCAAAAAAGGTCT <u>CGGGATCC</u> -ATGATTAAATCAAAAAAGGTCTAAACC CCCGCTCGAG-ATTATGATAGCGGCC	NdeI-NcoI BamHI XhoI
ORF 23	Forward Reverse	<u>CGCGGATCCCATATG</u> -GATGTTTCTGTTTCAGAC CCCGCTCGAG-TTTAAACCGATAGGTAAACG	BamHI-NdeI XhoI
ORF 24	Forward Forward Reverse	GGAATTC <u>CATATGGCCATGG</u> -TGATGCCGGAATGGTG <u>CGGGATCC</u> -ATGATGCCGGAATGGTG CCCGCTCGAG-TGTCAGCGTGCGCA	NdeI-NcoI BamHI XhoI
ORF 25	Forward Reverse	<u>GCGGATCCCATATG</u> -TATCGCAAAGTATTGC CCCGCTCGAG-ATCGATGGAATAGCCG	BamHI-NdeI XhoI
ORF 26	Forward Reverse	<u>GCGGATCCCATATG</u> -CAGCTGATCGACTATTC CCCGCTCGAG-GACATCGGCGGTTTT	BamHI-NdeI XhoI
ORF 27	Forward Forward Reverse	GGAATTC <u>CATATGGCCATGG</u> -AGACCTATTCTGTTTA <u>CGGGATCC</u> - CAGACCTATTCTGTTTATTTTAATC CCCGCTCGAG-GGGTTCGATTAAATAACCAT	NdeI-NcoI BamHI XhoI
ORF 28	Forward Forward Reverse	GGAATTC <u>CATATGGCCATGG</u> -ACGGCTGTACGTTGATGT <u>CGGGATCC</u> -AACGGCTGTACGTTGATG CCCGCTCGAG-TTTGTCAGAGGAATTCGCG	NdeI-NcoI BamHI XhoI
ORF 29	Forward Forward Reverse	<u>GCGGATCCCATATG</u> -AACGGTTTGATGCCCCG <u>CGCGGATCCGCTAGC</u> -AACGGTTTGATGCCCCG CCCGCTCGAG-TTTGTCTAAGTTCCTGATATG	BamHI-NdeI BamHI-NheI XhoI
ORF 32	Forward Reverse	<u>CGCGGATCCCATATG</u> -AATACTCCTCCTTTTG CCCGCTCGAG-GCGTATTTTTTGATGCTTTG	BamHI-NdeI XhoI
ORF 33	Forward Reverse	<u>GCGGATCCCATATG</u> -ATTGATAGGGATCGTATG CCCGCTCGAG-TTGATCTTTCAAACGGCC	BamHI-NdeI XhoI

<b>ORF 35</b>	Forward Forward Reverse	GCGGATCCCATATG-TTCAGAGCTCAGCTT CGCGGATCCGCTAGC-TTCAGAGCTCAGCTT CCCGCTCGAG-AAACAGCCATTTGAGCGA	BamHI-NdeI BamHI-NheI XhoI
<b>ORF 37</b>	Forward Reverse	GCGGATCCCATATG-GATGACGTATCGGATTTT CCCGCTCGAG-ATAGCCCGCTTTCAGG	BamHI-NdeI XhoI
<b>ORF 58</b>	Forward Reverse	CGCGGATCCGCTAGC-TCCGAACGCGAGTGGAT CCCGCTCGAG-AGCATTGTCCAAGGGGAC	BamHI-NheI XhoI
<b>ORF 65</b>	Forward  Forward Reverse	GGAATTCCATATGGCCATGG -TGCTGTATCTGAATCAAG  CGGGATCC-TTGCTGTATCTGAATCAAGG CCCGCTCGAG-CCGCATCGGCAGACA	  NdeI-NcoI BamHI XhoI
<b>ORF 66</b>	Forward Reverse	GCGGATCCCATATG-TACGCATTTACCGCCG CCCGCTCGAG-TGGATTTTGAGAGATGG	BamHI-NdeI XhoI
<b>ORF 72</b>	Forward Reverse	CGCGGATCCCATATG- AATGCAGTAAAAATATCTGA CCCGCTCGAG-GCCTGAGACCTTTGCAA	BamHI-NdeI XhoI
<b>ORF 73</b>	Forward Reverse	GCGGATCCCATATG-AGATTTTTCGGTATCGG CCCGCTCGAG-TTCATCTTTTTCATGTTTCG	BamHI-NdeI XhoI
<b>ORF 75</b>	Forward Reverse	GCGGATCCCATATG- TCTGTCTTTCAAACGGC CCCGCTCGAG-TTTGTTTTTGCAAGACAG	BamHI-NdeI XhoI
<b>ORF 76</b>	Forward Reverse	GATCAGCTAGCCATATG-AAACAGAAAAAACCGC CGGGATCC-TTACGGTTTGACACCGTT	NheI-NdeI BamHI
<b>ORF 79</b>	Forward Reverse	CGCGGATCCCATATG-GTTTCCGCCGCCG CCCGCTCGAG-GTGCTGATGCGCTTCG	BamHI-NdeI XhoI
<b>ORF 83</b>	Forward Reverse	GCGGATCCCATATG-AAAACCCTGCTGCTGC CCCGCTCGAG-GCCGCCTTTGCGGC	BamHI-NdeI XhoI
<b>ORF 84</b>	Forward Reverse	GCGGATCCCATATG-GCAGAGATCTGTTTG CCCGCTCGAG-GTTTGCCGATCCGACCA	BamHI-NdeI XhoI
<b>ORF 85</b>	Forward Reverse	CGCGGATCCCATATG- GCGGTTTGGGGCGGA CCCGCTCGAG-TCGGCGCGGCGGGC	BamHI-NdeI XhoI
<b>ORF 89</b>	Forward Forward Reverse	GGAATTCCATATGGCCATGG-CCATACCTTCTTATCA CGGGATCC-GCCATACCTTCTTATCAGAG CCCGCTCGAG-TTTTTTGCGATTAGAAAAAGC	NdeI-NcoI BamHI XhoI
<b>ORF 97</b>	Forward	GCGGATCCCATATG-CATCCTGCCAGCGAAC	BamHI-NdeI

<b>ORF 98</b>	Reverse	CCCGCTCGAG-TTCGCCTACGGTTTTTTG	XhoI
	Forward Reverse	GCGGATCCCATATG-ACGGTAACTGCGG CCCGCTCGAG-TTGTTGTTTCGGGCAAATC	BamHI-NdeI XhoI
<b>ORF 100</b>	Forward Reverse	GCGGATCCCATATG-TCGGGCATTTACACCG CCCGCTCGAG-ACGGGTTTCGGCGGAA	BamHI-NdeI XhoI
<b>ORF 101</b>	Forward Reverse	GCGGATCCCATATG-ATTTATCAAAGAAACCTC CCCGCTCGAG-TTTTCGCCTTTCAATGT	BamHI-NdeI XhoI
<b>ORF 102</b>	Forward Reverse	GCGGATCCCATATG-GCAGGGCTGTTTTACC CCCGCTCGAG-AAACGGTTTGAACACGAC	BamHI-NdeI XhoI
<b>ORF 103</b>	Forward Reverse	GCGGATCCCATATG-AACCACGACATCAC CCCGCTCGAG-CAGCCACAGGACGGC	BamHI-NdeI XhoI
<b>ORF 104</b>	Forward Reverse	GCGGATCCCATATG-ACGTGGGGAACGC CCCGCTCGAG-GCGGCGTTTGAACGGC	BamHI-NdeI XhoI
<b>ORF 105</b>	Forward Reverse	GCGGATCCCATATG-ACCAAATTTCAAACCCCTC CCCGCTCGAG-TAAACGAATGCCGTCCAG	BamHI-NdeI XhoI
<b>ORF 106</b>	Forward Reverse	GCGGATCCCATATG-AGGATAACCGACGGCG CCCGCTCGAG-TTTGTTCCCGATGATGTT	BamHI-NdeI XhoI
<b>ORF 109</b>	Forward Reverse	GCGGATCCCATATG-GAAGATTTATATATAATACTCG CCCGCTCGAG-ATCAGCTTCGAACCGAAG	BamHI-NdeI XhoI
<b>ORF110</b>	Forward Reverse	AAAGAATTC-ATGAGTAAATCCCGTAGATCTCCC AAACTGCAG-GGAAAACCACATCCGCACTCTGCC	EcoRI PstI
<b>ORF111</b>	Forward Reverse	AAAGAATTC-GCACC GCAAAGGCAAAAACCGCA AAACTGCAG-TCTGCGCGTTTTCGGGCAGGGTGG	EcoRI PstI
<b>ORF113</b>	Forward Reverse	AAAGAATTC-ATGAACAAAACCTCTATCGTGTGATTTTCAACCG AAACTGCAG-TTACGAATGCCTGCTTGCTCGACCGTACTG	EcoRI PstI
<b>ORF115</b>	Forward Reverse	AAAGAATTC-TTGCTTGTGCAACAGAAAAAGACGG AAAAAGTCGAC-CTATTTTTAGGGGC TTTTGC TTGTTGAAAAGCCTGCC	EcoRI SalI
<b>ORF119</b>	Forward Reverse	AAAGAATTC-TACAACATGTATCAGGAAAACCAATACCG AAACTGCAG-TTATGAAAACAGGCGCAGGGCGGTTTTGCC	EcoRI PstI
<b>ORF120</b>	Forward Reverse	AAAGAATTC-GCAAGGCTACCCCAATCCGCCGTG AAACTGCAG-CGGTTTGGCTGCCTGGCCGTTGAT	EcoRI PstI
<b>ORF121</b>	Forward Reverse	AAAGAATTC-GCCTTGGTCTGGCTGGTTTTTCGC AAACTGCAG-TCATCCGCCACCCACCTCGGCCATCCATC	EcoRI PstI

<b>ORF122</b>	Forward Reverse	AAAAAAGTCGAC-ATGTC TTACCGCGCAAGCAGTTC TCC AAACTGCAG-TCAGGAACACAAACGATGACGAATATCCGTATC	SalI PstI
<b>ORF125</b>	Forward Reverse	AAAGAATTC-GCGCTGTTTTTTGCGGCGGCGTAT AAACTGCAG-CGCCGTTTCAAGACGAAAAAGTCG	EcoRI PstI
<b>ORF126</b>	Forward Reverse	AAAGAATTC-GCGGAAACGGTCGAAG AAACTGCAG-TTAATCTTGTCTTCCGATATAC	EcoRI PstI
<b>ORF127</b>	Forward Reverse	AAAGAATTC-ATGACTGATAATCGGGGGTTTACG AAAAAAGTCGAC-CTTAAGTAACTTGCAAGTCCTTATC	EcoRI SalI
<b>ORF128</b>	Forward Reverse	AAAGAATTC-ATGCAAGCTGTCCGCTACAGGCC AAACTGCAG-CTATTGCAATGCGCCGCCGCGGAATGTTGAGCAGGCG	EcoRI PstI
<b>ORF129</b>	Forward Reverse	AAAGAATTC-ATGGATTTTCGTTTTGACATTATTTACGAATACCG AAACTGCAG-TTATTTTTTGATGAAATTTTGGGGCGG	EcoRI PstI
<b>ORF130</b>	Forward Reverse	AAAGAATTC-GCAGTACTTGCCATTCTCGGTGCG AAACTGCAG-CTCCGGATCGTCTGTAAACGCATT	EcoRI PstI
<b>ORF 131</b>	Forward Reverse	GCGGATCCCATATG-GAAATTCGGGCAATAAAAT CCCGCTCGAG-CCAGCGGACGCGTTC	BamHI-NdeI XhoI
<b>ORF 132</b>	Forward Reverse	GCGGATCCCATATG-AAAGAAGCGGGGTTTG CCCGCTCGAG-CCAATCTGCCAGCCGT	BamHI-NdeI XhoI
<b>ORF 133</b>	Forward Reverse	CGCGGATCCCATATG-GAAGATGCAGGGCGCG CCCGCTCGAG-AAACTTGTAGCTCATCGT	BamHI-NdeI XhoI
<b>ORF 134</b>	Forward Reverse	GCGGATCCCATATG-TCTGTGCAAGCAGTATTG CCCGCTCGAG-ATCCTGTGCCAATGCG	BamHI-NdeI XhoI
<b>ORF 135</b>	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAAAAGCTTT CCCGCTCGAG-AAATACCGCTGAGGATG	BamHI-NdeI XhoI
<b>ORF 136</b>	Forward Reverse	CGCGGATCCGCTAGC-ATGAAGCGGCGTATAGCC CCCGCTCGAG-TTCCGAATATTTGGAACTTTT	BamHI-NheI XhoI
<b>ORF 137</b>	Forward Reverse	CGCGGATCCCATATG-GGCACGGCGGGAAATA CCCGCTCGAG-ATAACGGTATGCCGCC	BamHI-NdeI XhoI
<b>ORF 138</b>	Forward Reverse	GCGGATCCCATATG-TTTCGTTTACAATTCAGGC CCCGCTCGAG-CGGCGTTTTATAGCGG	BamHI-NdeI XhoI
<b>ORF 139</b>	Forward Reverse	GCGGATCCCATATG-GCTTTTTTGGCGGTAATG CCCGCTCGAG-TAACGTTTCCGTGCGTTT	BamHI-NdeI XhoI

<b>ORF 140</b>	Forward Reverse	GCGGATCCCATATG-TTGCCACAGGCAGC CCCGCTCGAG-GACGATGGCAAACAGC	BamHI-NdeI XhoI
<b>ORF 141</b>	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAGCAGTCT CCCGCTCGAG-ATCTGTTGTTTTAAATATT	BamHI-NdeI XhoI
<b>ORF 142</b>	Forward Reverse	GCGGATCCCATATG-GATAATTCTGGTAGTGAAG CCCGCTCGAG-AAACGTATAGCCTACCT	BamHI-NdeI XhoI
<b>ORF 143</b>	Forward Reverse	GCGGATCCCATATG-GATACCGCTTTGAACCT CCCGCTCGAG-AATGGCTTCCGCAATATG	BamHI-NdeI XhoI
<b>ORF 144</b>	Forward Reverse	GCGGATCCCATATG-ACCTTTTTACAACGTTTGC CCCGCTCGAG-AGATTGTTGTTGTTTTTCG	BamHI-NdeI XhoI
<b>ORF 147</b>	Forward Reverse	GCGGATCCCATATG-TCTGTCTTTCAAACGGC CCCGCTCGAG-TTTGTTTTTGCAAGACAG	BamHI-NdeI XhoI

NB:

- restriction sites are underlined
- for ORFs 110-130, where the ORF itself carries an *EcoRI* site (eg. ORF122), a *SaII* site was used in the forward primer instead. Similarly, where the ORF carries a *PstI* site (eg. ORFs 115 and 127), a *SaII* site was used in the reverse primer.

TABLE II – Summary of cloning, expression and purification

ORF	PCR/cloning	His-fusion expression	GST-fusion expression	Purification
orf 1	+	+	+	His-fusion
orf 2	+	+	+	GST-fusion
orf 2.1	+	n.d.	+	GST-fusion
orf 4	+	+	+	His-fusion
orf 5	+	n.d.	+	GST-fusion
orf 6	+	+	+	GST-fusion
orf 7	+	+	+	GST-fusion
orf 8	+	n.d.	n.d.	
orf 9	+	+	+	GST-fusion
orf 10	+	n.d.	n.d.	
orf 11	+	n.d.	n.d.	
orf 13	+	n.d.	+	GST-fusion
orf 15	+	+	+	GST-fusion
orf 17	+	n.d.	n.d.	
orf 18	+	n.d.	n.d.	
orf 19	+	n.d.	n.d.	
orf 20	+	n.d.	n.d.	
orf 22	+	+	+	GST-fusion
orf 23	+	+	+	His-fusion
orf 24	+	n.d.	n.d.	
orf 25	+	+	+	His-fusion
orf 26	+	n.d.	n.d.	
orf 27	+	+	+	GST-fusion
orf 28	+	+	+	GST-fusion
orf 29	+	n.d.	n.d.	
orf 32	+	+	+	His-fusion
orf 33	+	n.d.	n.d.	
orf 35	+	n.d.	n.d.	
orf 37	+	+	+	GST-fusion
orf 58	+	n.d.	n.d.	
orf 65	+	n.d.	n.d.	
orf 66	+	n.d.	n.d.	
orf 72	+	+	n.d.	His-fusion
orf 73	+	n.d.	+	n.d.
orf 75	+	n.d.	n.d.	
orf 76	+	+	n.d.	His-fusion
orf 79	+	+	n.d.	His-fusion
orf 83	+	n.d.	+	n.d.
orf 84	+	n.d.	n.d.	

orf 85	+	n.d.	+	GST-fusion
orf 89	+	n.d.	+	GST-fusion
orf 97	+	+	+	GST-fusion
orf 98	+	n.d.	n.d.	
orf 100	+	n.d.	n.d.	
orf 101	+	n.d.	n.d.	
orf 102	+	n.d.	n.d.	
orf 103	+	n.d.	n.d.	
orf 104	+	n.d.	n.d.	
orf 105	+	n.d.	n.d.	
orf 106	+	+	+	His-fusion
orf 109	+	n.d.	n.d.	
orf 110	+	n.d.	n.d.	
orf 111	+	+	n.d.	His-fusion
orf 113	+	+	n.d.	His-fusion
orf 115	n.d.	n.d.	n.d.	
orf 119	+	+	n.d.	His-fusion
orf 120	+	+	n.d.	His-fusion
orf 121	+	n.d.	n.d.	
orf 122	+	+	n.d.	His-fusion
orf 125	+	+	n.d.	His-fusion
orf 126	+	+	n.d.	His-fusion
orf 127	+	+	n.d.	His-fusion
orf 128	+	n.d.	n.d.	
orf 129	+	+	n.d.	His-fusion
orf 130	+	n.d.	n.d.	
orf 131	+	+	+	n.d.
orf 132	+	+	+	His-fusion
orf 133	+	n.d.	+	GST-fusion
orf 134	+	n.d.	n.d.	
orf 135	+	n.d.	n.d.	
orf 136	+	n.d.	n.d.	
orf 137	+	n.d.	+	GST-fusion
orf 138	+	n.d.	+	GST-fusion
orf 139	+	n.d.	n.d.	
orf 140	+	n.d.	n.d.	
orf 141	+	n.d.	n.d.	
orf 142	+	n.d.	n.d.	
orf 143	+	n.d.	n.d.	
orf 144	+	n.d.	+	n.d.
orf 147	+	n.d.	n.d.	

**CLAIMS**

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, and 8.
2. A nucleic acid molecule which encodes a protein according to claim 1.
- 5 3. A nucleic acid molecule according to claim 2, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, and 7.
4. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..
- 15 5. A protein having 50% or greater sequence identity to a protein according to claim 4.
- 20
- 25
- 30



6. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..
7. An antibody which binds to a protein according to any one of claims 4 to 6.
8. A nucleic acid molecule which encodes a protein according to any one of claims 4 to 6.
9. A nucleic acid molecule according to claim 8, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289,

291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329,  
331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369,  
371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409,  
411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449,  
5 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489,  
491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529,  
531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569,  
571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609,  
611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649,  
10 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689,  
691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729,  
731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769,  
771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809,  
811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849,  
15 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889,  
& 891..

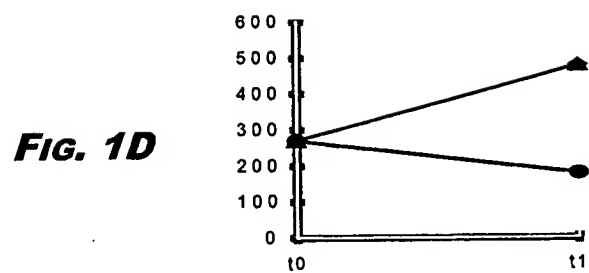
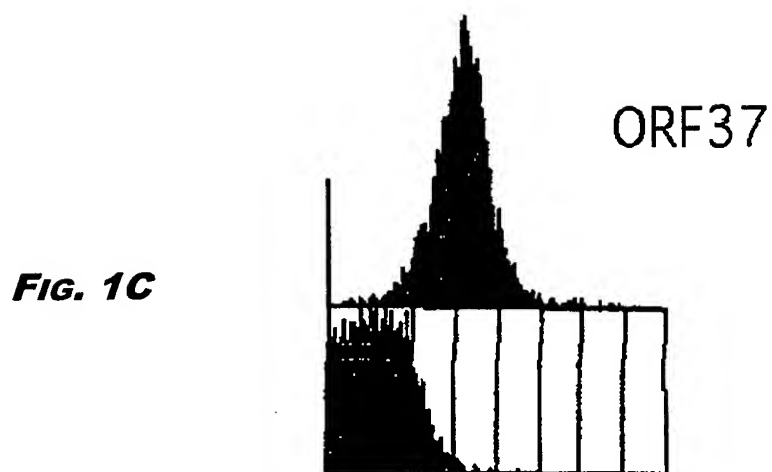
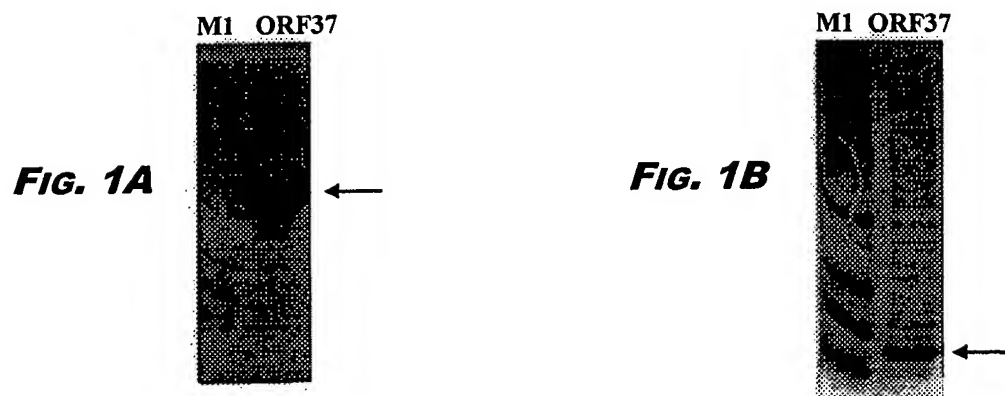
10. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the  
group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39,  
41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91,  
20 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133,  
135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173,  
175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213,  
215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253,  
255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293,  
25 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333,  
335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373,  
375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413,  
415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453,  
455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493,  
30 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533,  
535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573,  
575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613,  
615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653,  
655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693,  
35 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733,

735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, & 891..

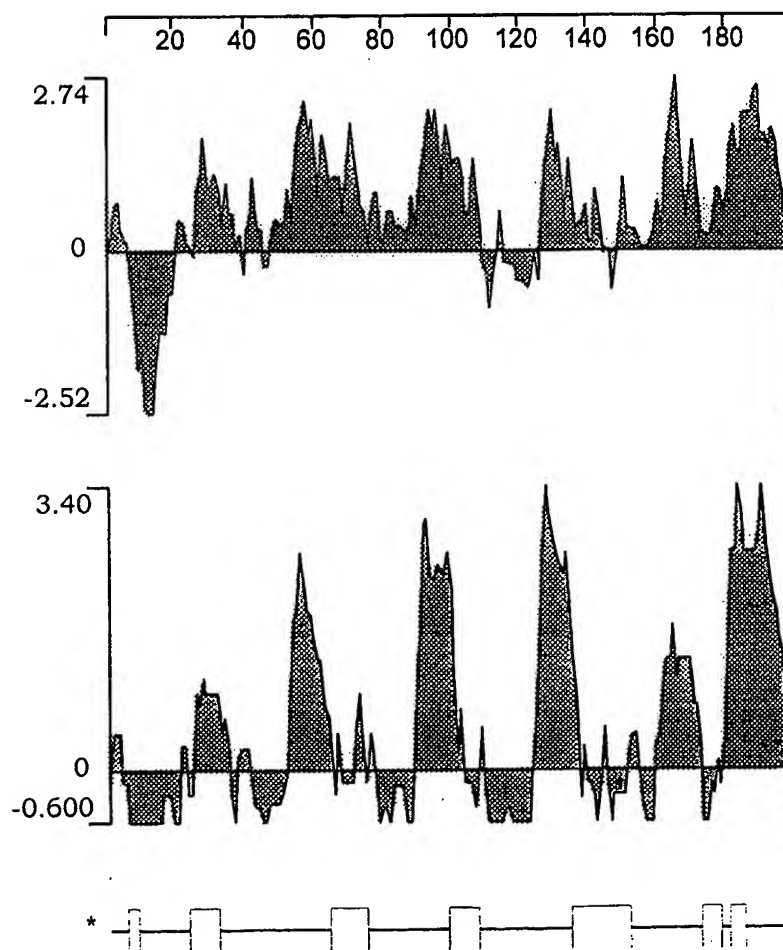
- 5 11. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 8 to 10.
12. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 8-11.
13. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any  
10 one of claims 8-12 under high stringency conditions.
14. A composition comprising a protein, a nucleic acid molecule, or an antibody according to any preceding claim.
15. A composition according to claim 14 being a vaccine composition or a diagnostic composition.
- 15 16. A composition according to claim 14 or claim 15 for use as a pharmaceutical.
17. The use of a composition according to claim 14 in the manufacture of a medicament for the treatment or prevention of infection due to Neisserial bacteria.

1/24

**FIGURE 1**



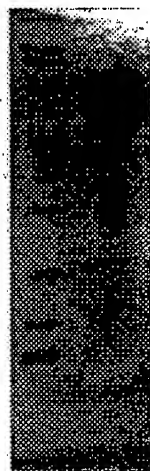
2/24

**Fig 1E**

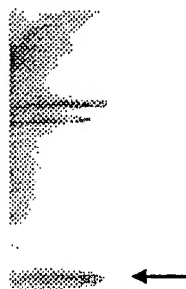
3/24

**FIGURE 2****Fig. 2A**

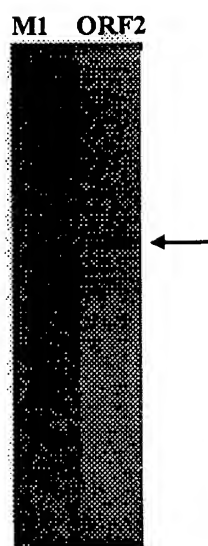
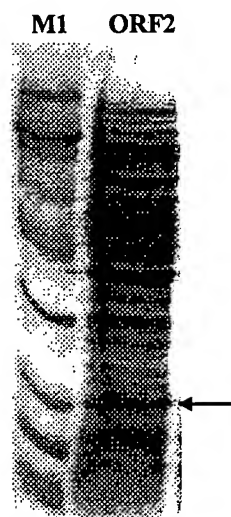
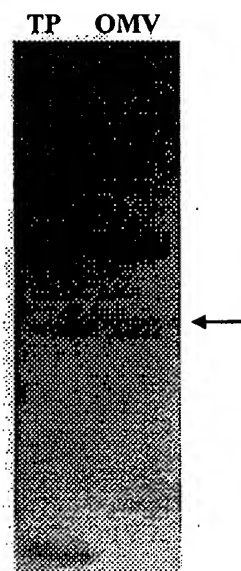
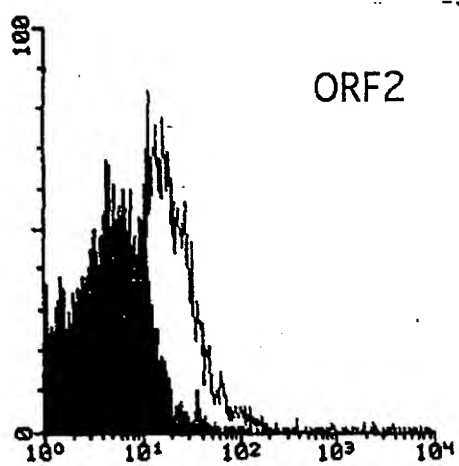
M1 ORF5

**Fig. 2B**

TP



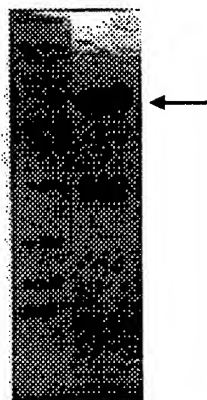
4/24

**FIGURE 3****FIG. 3A****FIG. 3B****FIG. 3C****FIG. 3D**

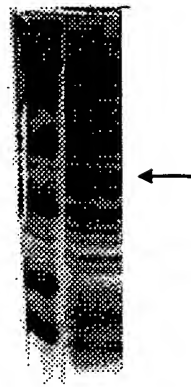
5/24

**FIGURE 4****Fig. 4A**

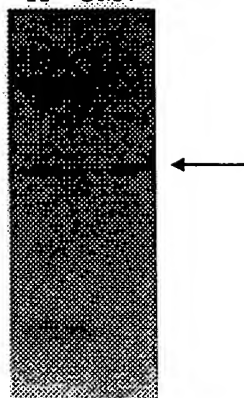
M1 ORF15

**Fig. 4B**

M2 ORF15

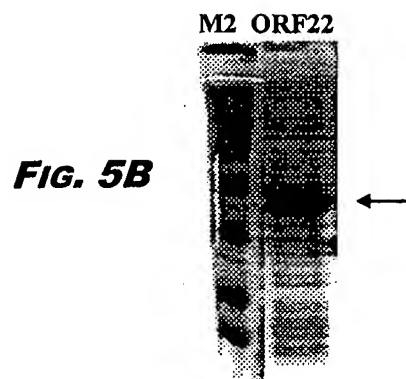
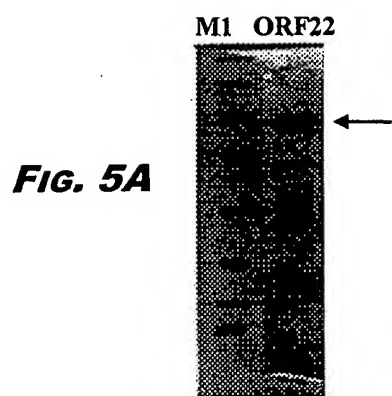
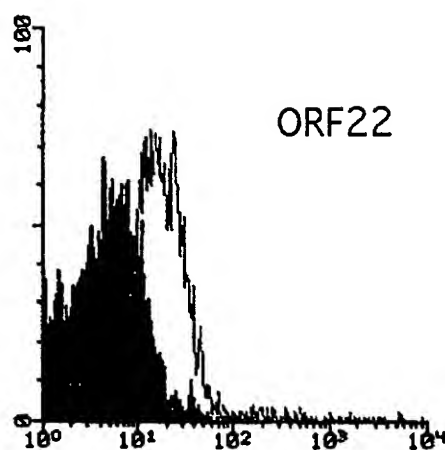
**Fig 4C**

TP OMV





6/24

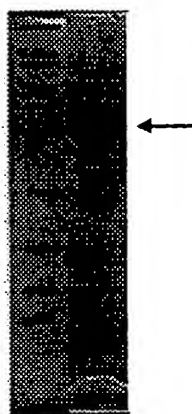
**FIGURE 5****Fig. 5C**

7/24

**FIGURE 6**

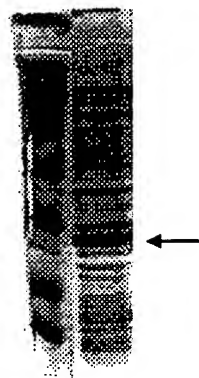
***Fig. 6A***

M1 ORF28



***Fig. 6B***

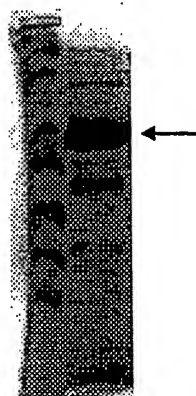
M2 ORF28



**FIGURE 7**

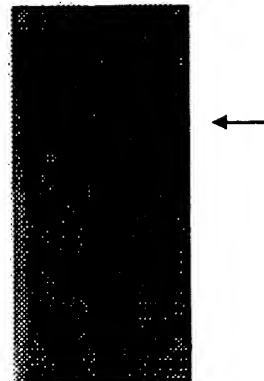
***Fig. 7A***

M1 ORF32



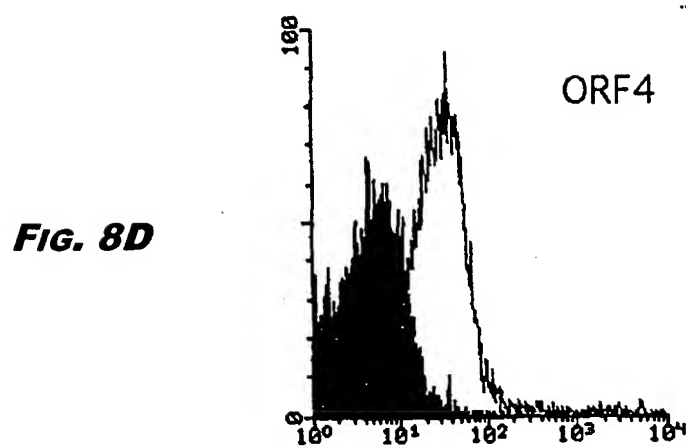
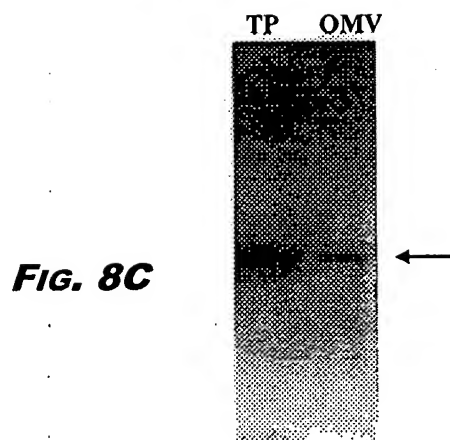
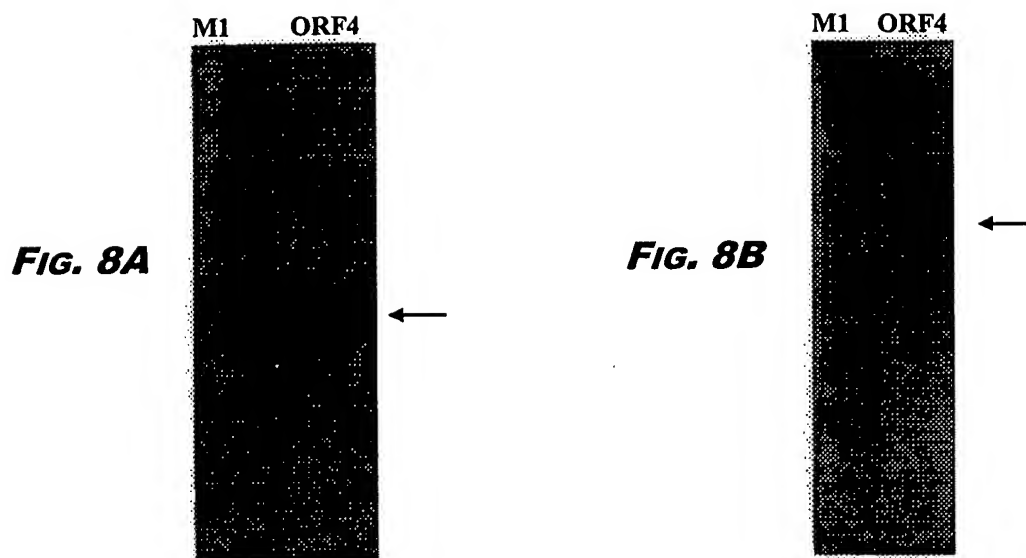
***Fig. 7B***

M1 ORF32

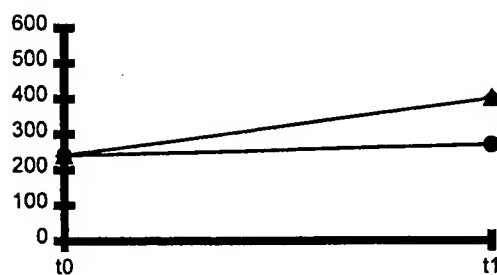
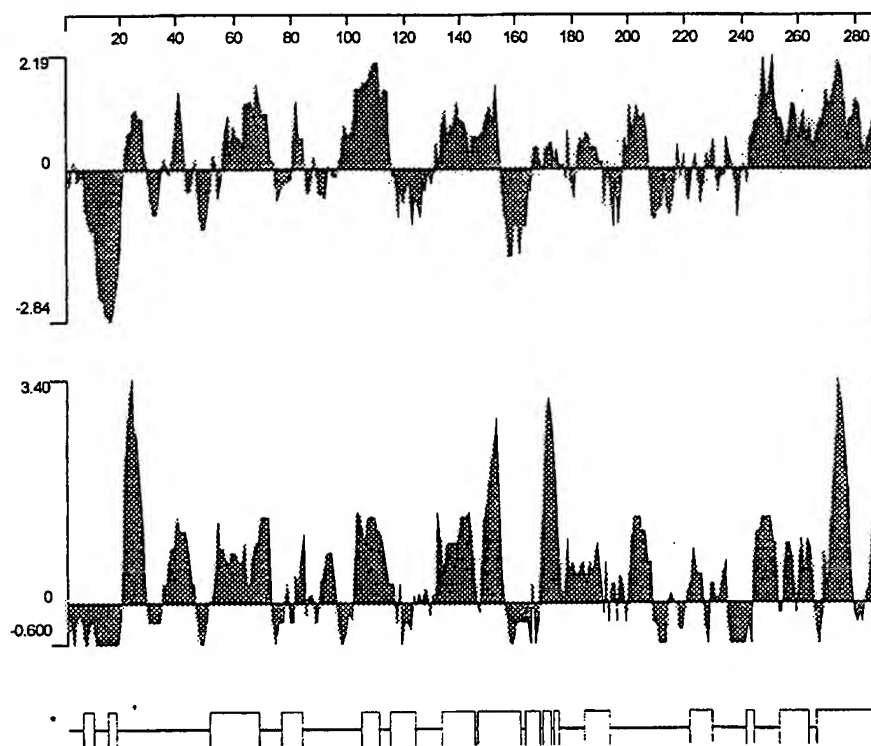


8/24

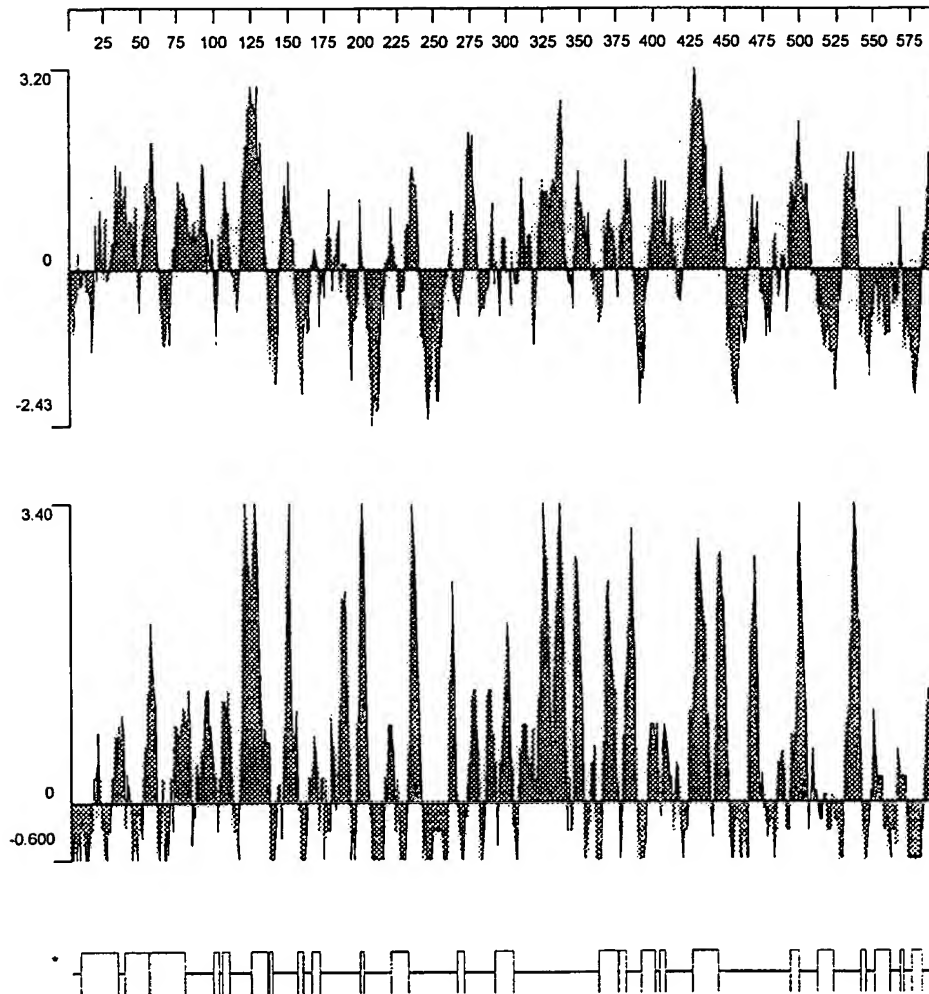
**FIGURE 8**



9/24

**FIG. 8E****FIG. 8F**

10/24

**FIGURE 9**

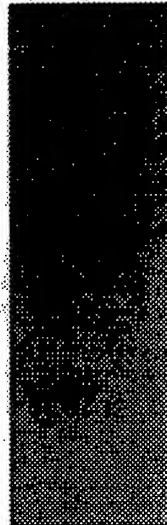
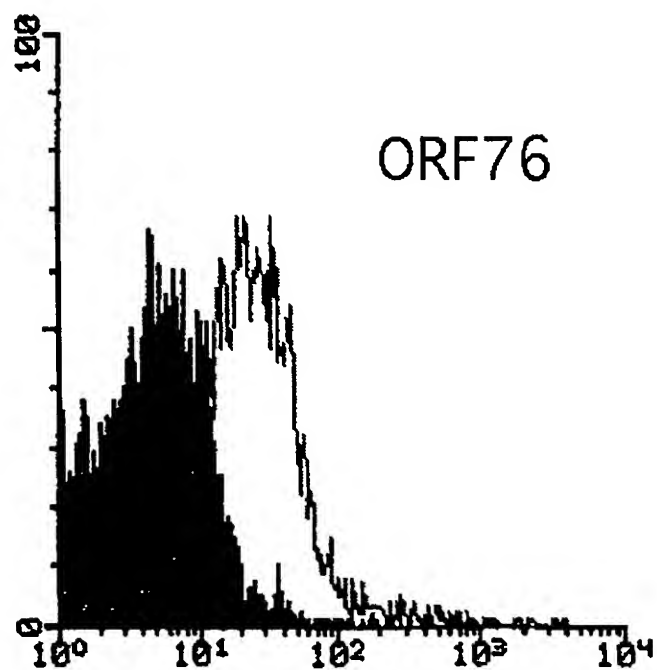
11/24

**FIGURE 10****FIG. 10A**

M1 ORF76

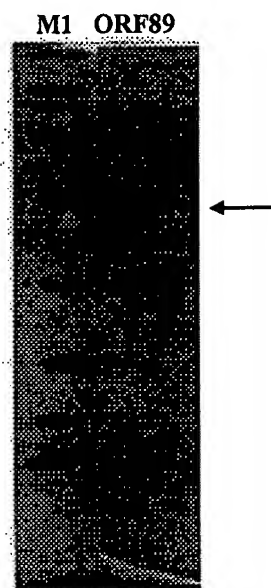
**FIG. 10B**

TP OMV

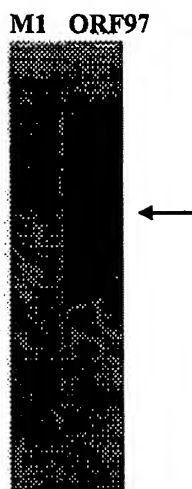
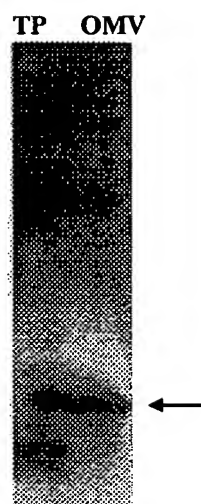
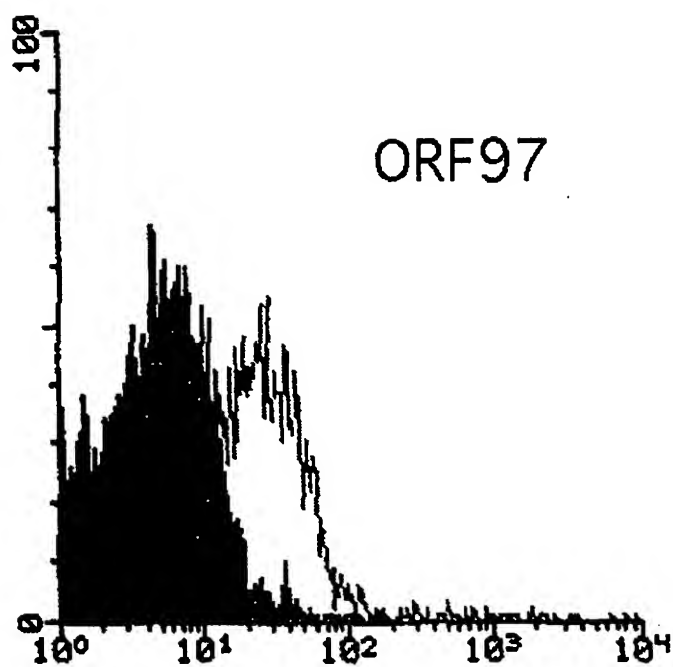
**FIG. 10C**

12/24

**FIGURE 11**

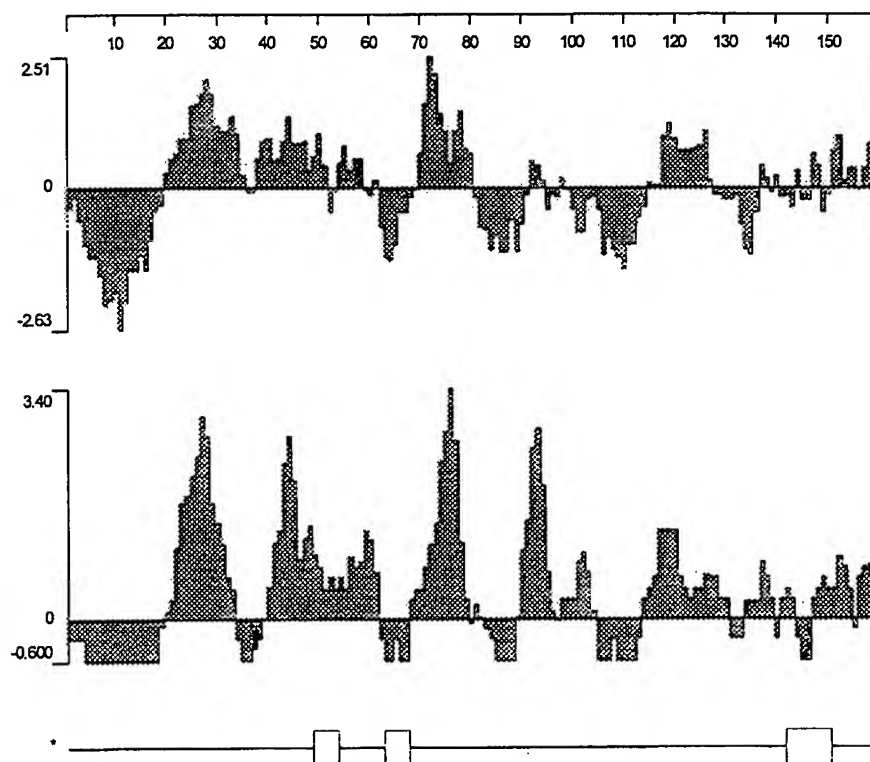


13/24

**FIGURE 12****FIG. 12A****FIG. 12B****FIG. 12C****FIG. 12D**



14/24

**FIG. 12E**

15/24

**FIGURE 13****Fig. 13A**

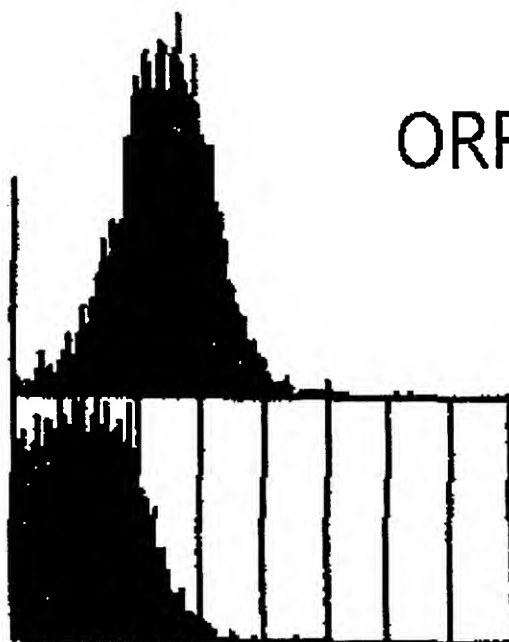
M1 ORF106

**Fig. 13B**

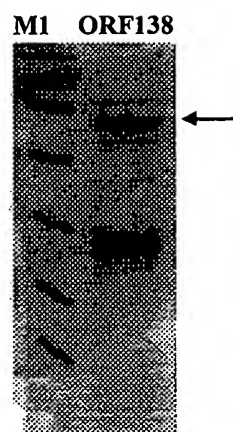
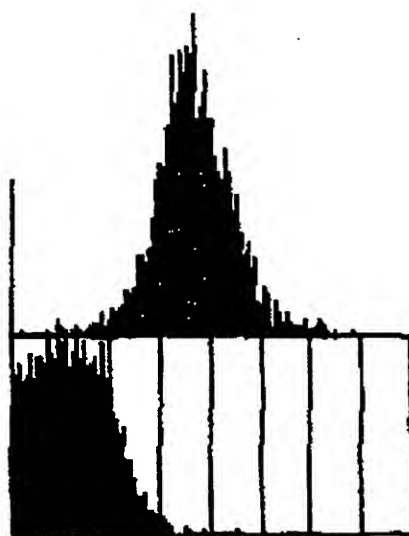
M2 ORF106

**Fig. 13C**

ORF 106



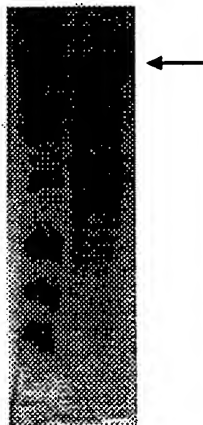
16/24

**FIGURE 14****FIG. 14A****FIG. 14B**

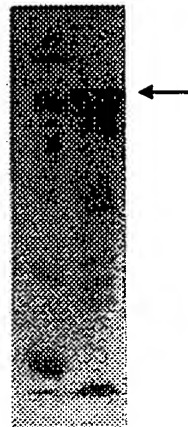
17/24

**FIGURE 15****Fig. 15A**

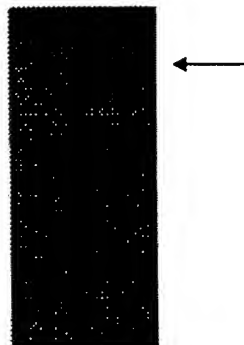
M1 ORF23

**Fig. 15B**

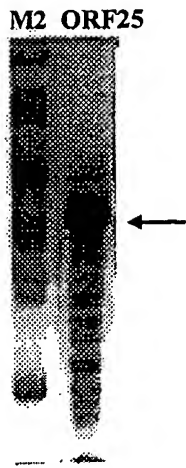
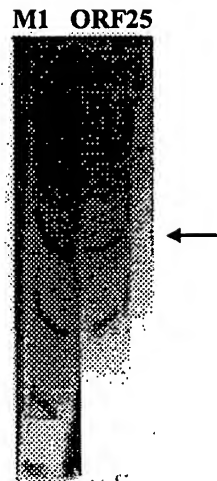
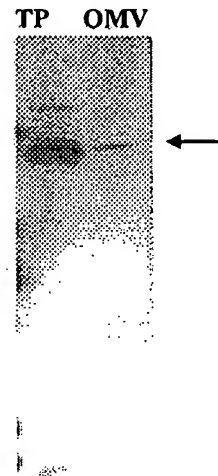
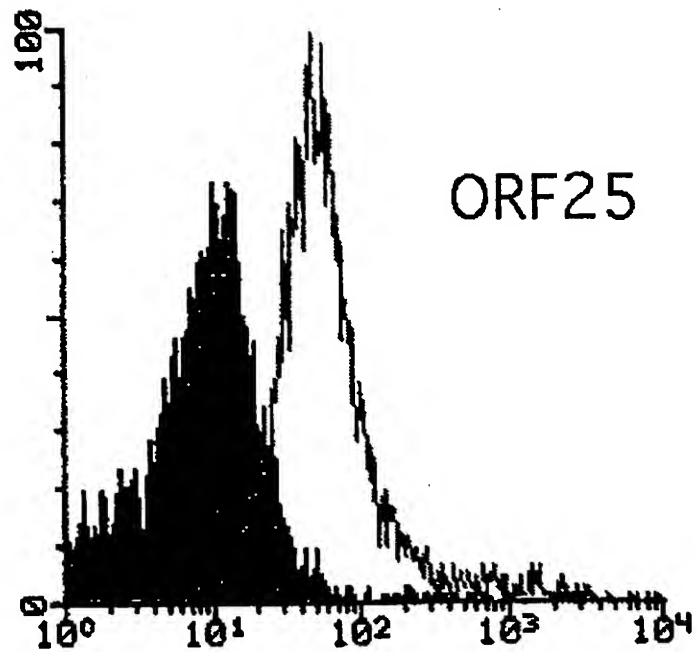
M2 ORF23

**Fig 15C**

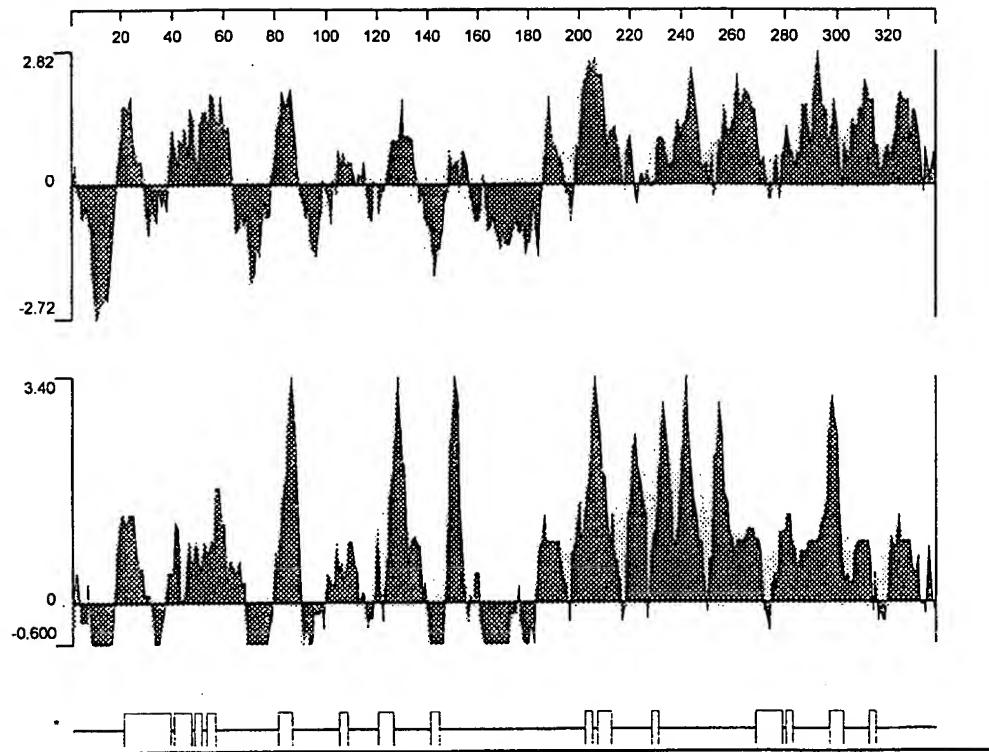
TP OMV



18/24

**FIGURE 16****Fig. 16A****Fig. 16B****Fig. 16C****Fig. 16D**

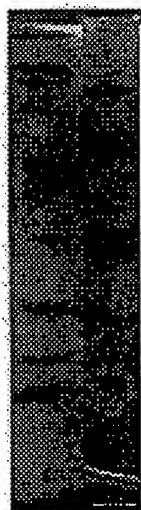
19/24

**FIG. 16E**

**FIGURE 17**

***FIG. 17A***

M1 ORF27

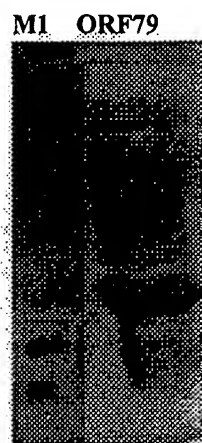
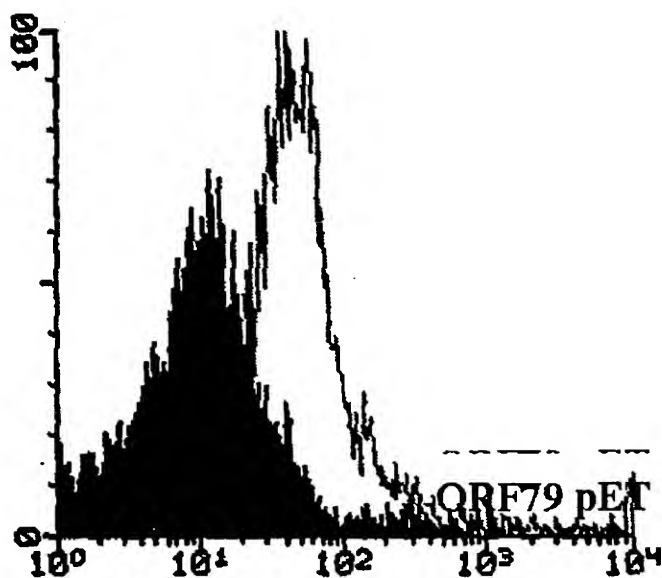


***FIG. 17B***

M2 ORF27

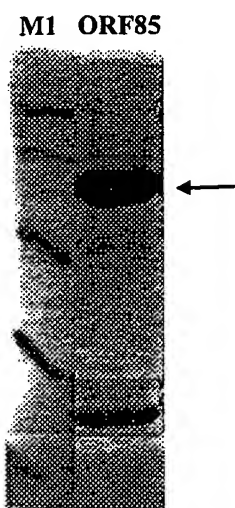
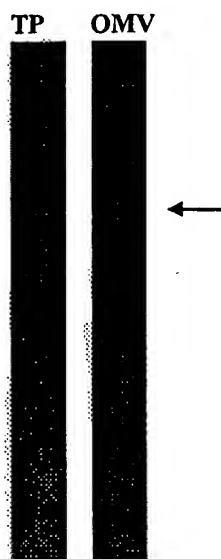
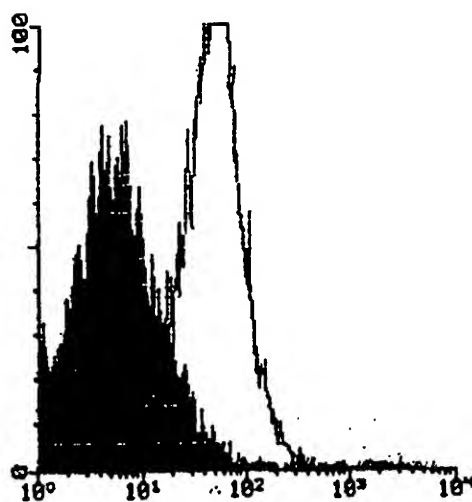


21/24

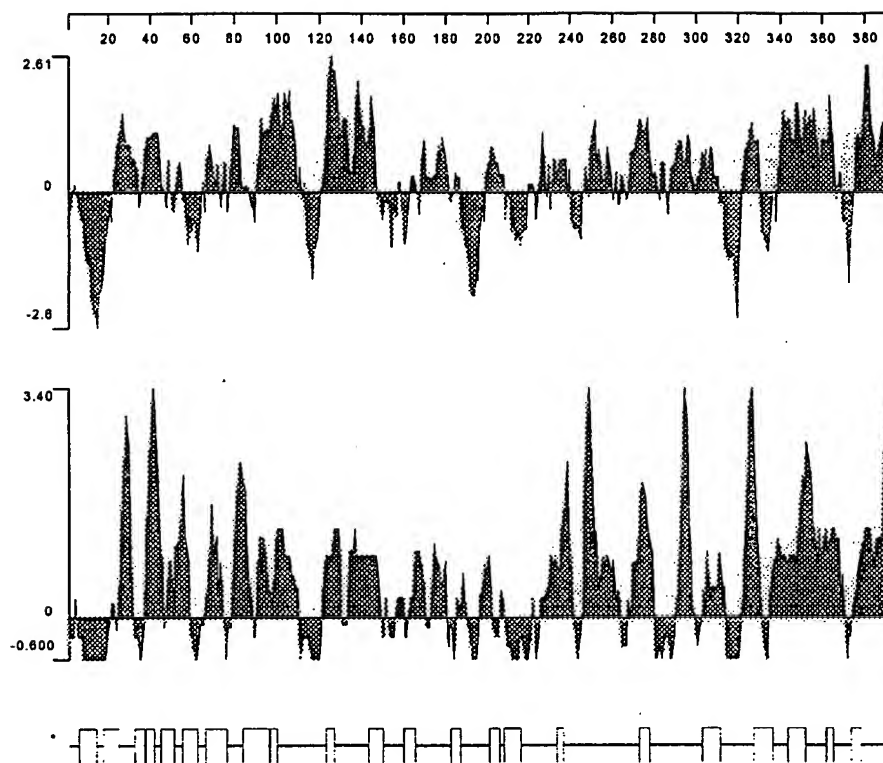
**FIGURE 18****FIG. 18A****FIG. 18B**



22/24

**FIGURE 19****FIG. 19A****FIG. 19B****FIG. 19C**

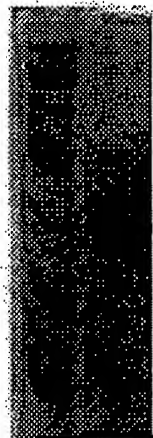
23/24

**Fig 19D**

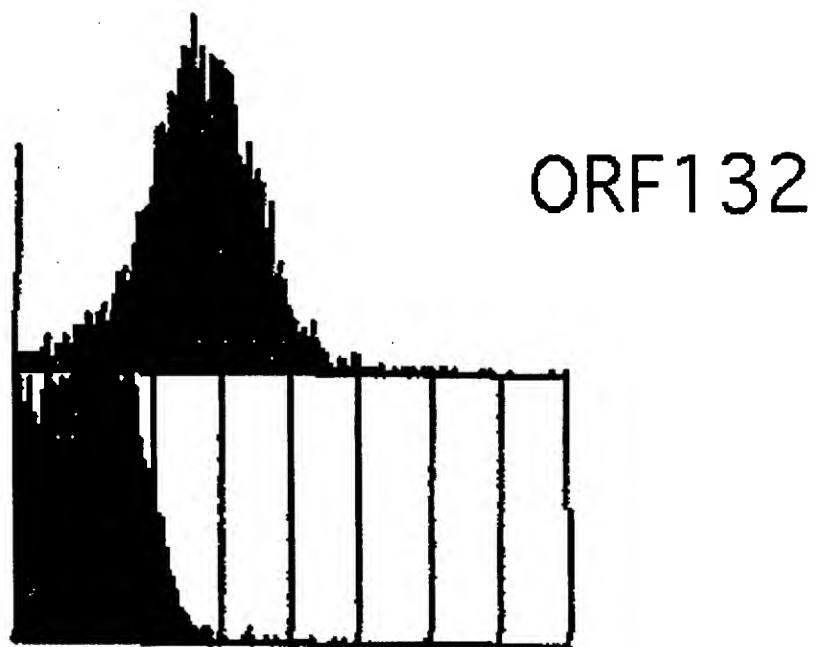
24/24

**FIGURE 20****FIG. 20A**

M1 ORF132

**FIG. 20B**

M2 ORF132

**FIG. 20C**

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☒ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**